Journal of Affective Disorders 239 (2018) 79–84

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Determinants of antidepressant response: Implications for practice and future clinical trials

Jonathan Rabinowitz[a,*], Nomi Werbeloff[b], Francine S. Mandel[c], Lauren Marangell[d], François Menard[e], Shitij Kapur[f]

[a] Bar Ilan University, Ramat Gan, Israel
[b] University College London, London, United Kingdom
[c] Pfizer, United States
[d] Eli Lilly, United States
[e] Lundbeck SAS, France
[f] University of Melbourne, Australia

A B S T R A C T

Background: Response to antidepressants in major depressive disorder is variable and determinants are not well understood or used to design clinical trials. We aimed to understand these determinants.

Methods: Supported by Innovative Medicines Initiative, as part of a large public-private collaboration (NEWMEDS), we assembled the largest dataset of individual patient level information from industry sponsored randomized placebo-controlled trials of antidepressant drugs in adults with MDD. We examined patient and trial-design-related determinants of outcome as measured by change on Hamilton Depression Scale or Montgomery–Asberg Depression Rating Scale in 34 placebo-controlled trials (drug, $n = 8260$; placebo, $n = 3957$).

Results: While it is conventional for trials to be 6–8 weeks long, drug-placebo differences were nearly the same at week 4 as at week 6 and with lower dropout rates. At the multivariate level, having any of these attributes was significantly associated with greater drug vs. placebo differences on symptom improvement: female, increasing proportion of patients on placebo, centers located outside of North America, centers with low placebo response (regardless of active treatment response) and using randomized withdrawal designs.

Limitations: Data on compounds that failed were not available to us. Findings may not be relevant for new mechanisms of action.

Conclusions: Proof of concept trials can be shorter and efficiency improved by selecting enriched populations based on clinical and demographic variables, ensuring adequate balance of placebo patients, and carefully selecting and monitoring centers. In addition to improving drug discovery, patient exposure to placebo and experimental treatments can be reduced.

1. Introduction

Antidepressants were first discovered in the 1950′s and in the late 1980′s serotonin reuptake inhibitors (SSRIs) were introduced following a large number of double-blind randomized placebo-controlled trials with different compounds. Most of these SSRI trials were six to eight weeks in duration without stratification. They all included adult patients with major depressive disorder, regardless of symptom profile – despite evidence that any or all of the following factors may affect clinical response; age, sex (Kornstein and McEnany, 2000; Khan et al., 2005), geographic region (Khin et al., 2011) (See studies in Supplementary table). In addition, the literature suggests that trials could be shorter (Rutherford et al., 2009; Tedeschini et al., 2011) and that removing centers with unrealistically high or low placebo response (blinded to active treatment response) (Mallinckrodt and Prucka, 2010; Merlo-Pich et al., 2010) could heighten placebo-active treatment differences. However, these measures to improve trials have not been adequately tested to confidently include them in clinical trials. To complicate matters further, nearly half the patients dropped out of these trials (Rutherford et al., 2013) and there are international differences in study results according to European Medicines Agency (2009) raising methodological questions. These findings, along

* Corresponding author.
E-mail address: jonathan.rabinowitz@biu.ac.il (J. Rabinowitz).

https://doi.org/10.1016/j.jad.2018.06.039
Received 13 April 2017; Received in revised form 13 June 2018; Accepted 14 June 2018
Available online 02 July 2018

with a fair number of negative or failed trials (where an established drug fails to separate from placebo) (Khin et al., 2012), the moderate superiority over placebo (Leucht et al., 2009) and the rising cost and difficulties of completing these trials, have led to questions whether the current approach to trials is most effective – and whether more focused and shorter trials might yield informative results especially in early phases of clinical development.

To address the above mentioned impediments to drug development, the National Advisory Mental Health Council (2010) has recommended sharing of data to improve efficiency and decrease cost of therapeutic development. This could enable identifying moderators and mediators of treatment effects, and facilitate establishing a biologically-based discovery process. In concert with this, as part of the European Union funded Innovative Medicines Initiative, an academic and industry collaboration, we merged individual patient data from 34 randomized controlled trials (RCT) from four pharmaceutical companies. We explored determinants of antidepressant response in major depressive disorder, optimal trial duration, and whether these findings could be used to design more efficient trials in general, and specifically proof of concept trials.

We examined which key demographic and clinical variables, as well as study design features, influenced response, and if so, in what way. Next, as treatment response may be reached earlier than six weeks, we tested if study conclusions could have been reached earlier. Finally, based on previous literature on the inflation of baseline scores stemming from enrollment pressures (e.g., DeBrota et al., 1999; Kobak et al., 2010), we examined whether patients who just met symptom inclusion criteria were overrepresented and whether this appeared to affect study results. We speculated that if there was an overrepresentation of patients just meeting inclusion criteria, this may suggest that scores may have been inflated for purposes of including them in the study (the so-called “baseline inflation”). We then examined whether their exclusion might have resulted in different conclusions.

2. Method

The NEWMEDS repository includes anonymized patient data from controlled studies to treat depression from the 39 randomized placebo-controlled trials (n = 12,217) (1983–2007) of citalopram, duloxetine, escitalopram, quetiapine and sertraline. This included all the acute placebo-controlled trials of major depressive disorder in non-enriched (e.g., no major psychiatric comorbidities) adult populations, sponsored or owned by Pfizer, Eli Lilly, AstraZeneca and Lundbeck. We examined patient and trial-design-related determinants of outcome as measured by change on the Hamilton Depression Scale or the Montgomery–Asberg Depression Rating Scale in 34 placebo-controlled trials (drug, n = 8260; placebo, n = 3957). Eight, out of 22 active-placebo studies, were negative studies, and 5/17 studies with active comparators were failed studies (no difference on study drug and active comparator vs. placebo). Five of 39 studies were relapse prevention studies with open label randomized withdrawal designs prior to randomization of responders.

Results of the individual studies (listed in Supplementary Table 1) have been publicized. These data have not been previously pooled into a single dataset. All drugs were grouped and compared to placebo. Each study had been approved by the relevant IRB when and where it was conducted. All studies included informed written consent of study participants. The first and second authors of this paper had full access to all the data in the studies, conducted all of the statistical analyses and take responsibility for the integrity of the data and the accuracy of the data analysis. There was no commercial funding for this work.

2.1. Measures

Studies used the MADRS (19 studies) or HAM-D (34 studies) and 14 studies used both. For combined analysis, we estimated the HAM-D based on the MADRS using equipercentile scaling (relative rank order within each measure). For randomized withdrawal designs, double-blind period baseline was used for change from baseline calculations.

2.2. Completeness of data

Complete data was available for all 12,217 subjects on sex, trial identifier, year of study and study arm. Data was missing on age for 15 subjects, on region for 434 subjects and site identifier for 6329 subjects. For purposes of analysis the 434 subjects with missing region were included in the “other” region group and age for the missing 15 subjects was replaced with mean age. Site was only included in one of set of analyses.

2.3. Analysis plan

Differential effects of key variables available at baseline on drug vs. placebo response were examined primarily based on the literature using a pre-specified analytic plan. The individual participant data from all studies were modeled simultaneously while accounting for the clustering of participants within studies as per the one-step approach to individual participant data meta-analysis, as described by Riley et al. (2010). Specifically, we conducted a multi-level Mixed Model Repeated Measures (MMRM) analyses with subjects nested within studies, controlling for baseline using scaled identity matrix. Patient level fixed effects studied were age (quartile), sex and treatment (drug vs. placebo). Study level-fixed effects were: investigated drug, region (Western Europe, Eastern Europe, North America), proportion of patients on placebo (25% or less, 26% to 35%, greater than 35%), design (standard vs. withdrawal), outcome measure (HAM-D or MADRS) and year of study. Adjusted marginal means, F test, degrees of freedom, p value and Cohen's d effect size score are reported. Effects of study year were examined by testing for linear effects in placebo drug difference using the MMRM estimated marginal means. Because site was not available for 52% of the subjects, site was not included as a level in the main analysis but studied in a second round of analysis directed at studying the effects of sites with vs. without unrealistically high placebo response.

A separate round of analysis was done of baseline inflation as only some studies had symptom level inclusion criteria. Given difficulties in recruiting, patients who are just below the eligibility threshold may have had their scores unintentionally inflated so that they may be included. These patients would be expected to show a more pronounced improvement early on in treatment (with both drug and placebo), and by increasing overall response in the placebo group may mitigate against finding a true difference. Specifically, twenty-eight studies had lower symptom level inclusion criteria (not including randomized withdrawal designs). These studies have screening data on 8990 patients. Of these, 395 patients (4.4%) had screening scores below the bottom inclusion criteria. These patients were removed from further analyses. To examine baseline inflation, subjects were grouped based on 5-point grouping of their baseline score from the bottom symptoms inclusion criteria. A potential baseline inflation was defined as patients with screening scores within 5 points of the bottom inclusion criteria (4,175 patients (48.6%) met this criteria and were the largest group (the next adjacent 5 point groups: 35.6%, 13.1%, 2.1% and 0.3%).

In addition, we tested to see whether shorter trials might be feasible. We examined percent of six-week difference between drug and placebo already discernible at each previous week. For example, if week 6 total difference between drug and placebo was five points and the week 5 difference was four points, then 80% (4/5) of this difference was discernible at week 5. Since in most trials a difference is considered to be statistically significant at a p-value of $< .05$, we examined if a drug-placebo difference which met this criteria at week six would also have met this criteria had the trial been stopped earlier (e.g., at three, four, or five weeks). All analyses were conducted using SPSS Version 23