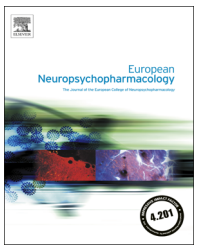
European Neuropsychopharmacology (2015) 25, 1999–2007

www.elsevier.com/locate/euroneuro



Usefulness of interim analyses in portending study results in antipsychotic and antidepressant trials

J. Rabinowitz[a,*], N. Werbeloff[a], F. Mandel[b], F. De Ridder[c], A. Schacht[d], F. Menard[e], I. Caears[c], V. Stauffer[d], S. Kapur[f]

[a]*Bar Ilan University, Ramat Gan, Israel*
[b]*Pfizer, New York, NY, USA*
[c]*Janssen Research & Development, Beerse, Belgium*
[d]*Eli Lilly, Indianapolis, IN, USA*
[e]*Lundbeck SAS, Paris, France*
[f]*Institute of Psychiatry, Kings College, London, United Kingdom*

Received 16 November 2014; received in revised form 14 July 2015; accepted 20 July 2015

KEYWORDS
Interim analysis;
Clinical trials

Abstract
It is unknown whether interim analyses portend final study results. Fatigue, pressure to complete trials and recruitment differences may mitigate against this. We examined the similarity of efficacy results of the first and second half of recruited patients to complete trials and explore possible intervening variables. Using data from the NewMeds repository of patient level data from placebo-controlled randomized trials of antipsychotics (AP) (22 studies, n=7056) and antidepressants (AD) (39 studies, n=12,217) we compared treatment effect size (placebo vs. active treatment) of the first and second half of patients recruited in completed trials. We found that in AP studies median difference in treatment effect between cohorts was −0.03, indicating that overall first and second cohorts yielded similar results. In AD studies, median difference between cohorts was 0.04, indicating that overall the second cohort had slightly larger active-placebo-difference. Overall, on average there were minimal differences in effect size between the first and the second cohorts, and in 30 of 39 trials interim results were a good estimate of the results on the 2nd cohort. In AD trials first and second cohort results were more similar when the proportion of patients per study centre and recruitment time of the two cohorts was similar. Results suggest that interim analyses in AD and AP studies may reliably serve to estimate ultimate effects and, at least in AD trials, are more accurate when the same sites are used to a similar extent and recruitment time of the two consequent cohorts is similar.

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

http://dx.doi.org/10.1016/j.euroneuro.2015.07.017
0924-977X/

1. Introduction

An objective of an interim analysis, conducted after a cohort of trial participants has completed the study, is to re-examine assumptions made about effect sizes in powering the study and to re-estimate sample size to see if additional patients might be needed. Yet, another is to portend the outcomes of a study and decide on whether it should be ended early for futility or, in some exceptional areas of medicine, for a resounding success. The interim analysis is predicated on the notion that the outcomes of the new cohort (i.e., post-interim analysis) will be similar to the original cohort (i.e., included in interim analysis). There is a risk that clinical trials that are stopped early for failure based on interim analysis may under estimate effect sizes and those that end early for success may overestimate effect sizes and in both of these scenarios to find implausible results (Pocock and Hughes, 1989). When interim analyses in randomized clinical trials (RCTs) identify larger than expected treatment effects, investigators may conclude, before completing the trial as planned, that one treatment is superior to the other. Such trials often receive much attention. Since the trials are truncated, the results that would have been obtained were the trial to run its course are unknown. Taking the interim analysis results at face value may be misleading if the decision to stop the trial or not to increase sample size resulted from catching the apparent benefit of treatment at a "random high." When this occurs, data from future trials would yield a more conservative estimate of treatment effect, the so-called "regression to the truth" effect (Montori et al., 2005). Because trials stopped based on an interim analysis end early, little is known about the accuracy of the interim result in estimating the treatment effect that would have been observed had the trial been completed. Aside from regression to the mean, there are trial related reasons that could cause non-trivial differences in results. For example, if the second cohort is drawn from a different subpopulation within the inclusion and exclusion criteria set by the protocol, having exhausted the supply of very symptomatic patients recruited in the first cohort, the second cohort consists of less symptomatic patients, this assumption might not be valid. Such differences can stem from slow recruiting sites and new centres added. Also trial fatigue may negatively affect results of the second cohort, i.e., there is concern that as the trial goes on enthusiasm wains as ease of recruiting, sustaining and maintaining qualified subjects in the study diminishes

The major objective of this paper is to examine the extent to which efficacy study results differ between the first and second half of patients recruited in antipsychotic and antidepressant studies. A secondary objective is to examine the possible role of changes of study centres over time and difference in duration of recruitment between the two cohorts.

2. Experimental procedures

Using data from the NEWMEDS repositories, one of placebo-controlled randomized trials of antipsychotic and the other placebo-controlled randomized trials of antidepressant medications, we examined the similarity of placebo-active differences between the first and second half of patients recruited into completed trials. NEWMEDS antipsychotics repository includes anonymized individual data from 29 placebo-controlled trials of second-generation antipsychotics (drug, $n=6971$, placebo, $n=2200$) conducted by AstraZeneca, Eli Lilly, Janssen, Lundbeck and Pfizer. Dates of study entry were available for 22/29 studies. Only those 22 studies were included in the analysis described (studies listed in online supplementary Table 1). NEWMEDS antidepressants repository includes anonymized individual data from 39 placebo-controlled trials (drug, $n=8260$, placebo, $n=3957$) conducted by AstraZeneca, Eli Lilly, Lundbeck and Pfizer (studies listed in online supplementary Table 2).

For the purpose of analyses all the active arms of a study were pooled together and compared to placebo. Patients in each study were divided into two cohorts at the point at which half of the patients had been randomized. The first cohort was regarded as the data contributing to the interim analysis. Placebo vs. active treatment differences on the PANSS in the antipsychotic trials and MADRS/HAMD in the depression trials were examined using ANCOVAs on change from baseline to endpoint (LOCF) controlling for baseline, for each study separately for each of the two cohorts. The difference between placebo vs. active treatment effects of the two cohorts were computed. Effect size (Cohen's d) was derived from the partial Eta Square produced by the ANCOVA [d=SQRT(Eta Square/(1−Eta Square))*2] (Cohen, 1988).

Effects of continuity of reliance on study centres between cohorts was also examined. To see if the consistency of mix of patients in each time cohort affects outcomes, we computed the proportion of patients per study centre for time 1 and time 2 and correlated the absolute difference in proportions with the absolute difference in effect sizes between the cohorts. For example, if for time 1 a given centre randomized 10 out of the 100 patients in the first cohort and 7 out of the next cohort of 100 patients, then for the first cohort they would have a value of 10% and of 7% for the second. The difference in the distribution of patients between cohorts for this site would then be 3%. Data on recruitment centre was available for 16 antipsychotic trials and 20 antidepressant trials. In addition, we computed the difference in recruitment time of the two cohorts (e.g., if the first cohort took 100 days to recruit and the second 200 days, the difference would be 100) and correlated this with the absolute difference in effect size. Spearman rank order correlations were used. We also examined homogeneity of the characteristics of patients in the trials by doing a pairwise comparison of demographic characteristics at baseline (for both AP and AD trials) and psychiatric history (available only for AP trials) between patients in the first and second cohort.

3. Results

In the antipsychotic studies (Table 1, Figure 1) the mean difference in active vs. placebo in effect size (weighted by sample size in the study) between cohorts was −0.01 and the median was −0.03, indicating that overall the first and second cohorts yielded similar results. In seven trials the difference was less than <-0.10, in 5 trials it was small ranging from −0.10 to +0.10; and 10 trials it was greater than +0.10. (A negative difference means that the first cohort showed a smaller effect size than second cohort and a positive, that the first cohort showed a larger effect size than the second cohort.). Overall, in 15 of 22 trials interim results either underestimated ultimate study effects and could have led to unnecessarily increasing sample sizes or would have been so small which would not have led to reducing sample sizes. Thus a reliance on the interim analysis would not have led to missing an efficacy signal.