Neurology, Psychiatry and Brain Research 30 (2018) 12–21

Contents lists available at ScienceDirect

Neurology, Psychiatry and Brain Research

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

Factors associated with placebo response in depression trials: A systematic review of published meta-analyses (1990–2017)

Joep H. Schoemaker[a,*], Sanja Kilian[b], Robin Emsley[b], Ad J.J.M. Vingerhoets[a]

[a] Tilburg University, Department of Medical & Clinical Psychology, PO Box 90153, 5000 LE Tilburg, Netherlands
[b] Stellenbosch University, Faculty of Medicine and Health Sciences, PO Box 241, Tygerberg Campus 8000, Cape Town, South Africa

ARTICLE INFO

Keywords:
Depression
Placebo effect
Systematic review

ABSTRACT

Background: Placebo response is common in patients with major depressive disorder (MDD) and decreases the likelihood of demonstrating drug superiority over placebo in a randomized, controlled trial (RCT). This paper aims to review the collective evidence for particular patient characteristics and trial features being associated with placebo response in MDD.
Methods: MEDLINE/PubMed publication database and Cochrane Library were searched for meta-analyses of placebo response in MDD, published in English from January 1990 to December 2017. The evidence for factors predicting a low or high placebo response was tabulated and weighted on the basis of methods, results, and quality of supporting studies.
Results: We identified 58 papers, examining the possible association of 40 different factors with placebo response in MDD. Research methods varied considerably across articles so that our reporting remained descriptive. The evidence for any factor being associated with placebo response in MDD appeared very weak to weak.
Limitations: Since none of the pooled analyses that we included could be regarded as a meta-analysis in its strict sense, and analytical approaches varied considerably, the current work is descriptive only, and without formal statistical analysis.
Conclusions: Despite 25 years of pooling data from RCTs in MDD, there is no single factor for which strong evidence exists that it influences placebo response.

1. Introduction

Sponsors of randomized controlled trials (RCTs) have been facing an increasing response rate to placebo in major depressive disorder (MDD) and other neuropsychiatric disorders over the last decades, resulting in failed studies, delayed or abandoned projects, and steep increases in Research and Development costs (Ackerman & Greenland, 2002; Kemp et al., 2010; Montgomery & Kasper, 2007; Sysko & Walsh, 2007; Welten et al., 2015). In the previous 25 years, this has led to a multitude of pooled analyses investigating predictors of placebo response in MDD.

The evidence from pooled analyses is frequently not convincing and sometimes even contradictory for the many predictors examined (Papakostas, Ostergaard, & Iovieno, 2015). This could be due to sampling bias (e.g., when factors are explored on multiple occasions for their association with placebo response under non-uniform conditions), or methodological flaws, such as the use of inappropriate statistical models, regression to the mean effects, and the use of the relative efficacy of antidepressants versus placebo as outcome variable (which can only provide indirect evidence for a factor being associated with placebo response). In order to identify moderators of placebo response, it is essential to use data from as many RCTs as possible, preferably at patient-level rather than study-level, and to correct for heterogeneity in study design when executing pooled analyses. The use of individual study participant data is ideal in any meta-analysis, in that it allows standardizing the statistical analysis of each study, obtaining summary results directly, checking the assumptions of models, examining interactions, and adjusting each patient's change score by their baseline value and other patient-level characteristics.

A limited number of meta-analyses have been reviewed by several authors (here called "meta-reviews"), with conclusions sometimes being drawn without regard to the methods applied in the meta-analyses, the existence of contradictory results reported elsewhere, or without providing criteria for weighing the level of evidence coming from these meta-analyses (Papakostas et al., 2015; Rutherford & Roose, 2013; Weimer, Colloca, & Enck, 2015). On the basis of a review of 13 meta-analyses, Rutherford and Roose concluded that there is "strong" evidence for a positive association between placebo response in depression RCTs and (*1*) a lower probability of receiving placebo or multiple active

* Corresponding author.
E-mail addresses: j.h.schoemaker@uvt.nl (J.H. Schoemaker), rae@sun.ac.za (R. Emsley).

https://doi.org/10.1016/j.npbr.2018.04.002
Received 23 February 2018; Received in revised form 29 March 2018; Accepted 4 April 2018

treatment arms, (*2*) the average number of study sites in a RCT, and (*3*) poor rater blinding, without providing the criteria on the basis of which they reached this conclusion. After reviewing 14 meta-analyses in depression (5 original, and 9 already reviewed by Rutherford and Roose), Weimer et al. reported (*1*) lower probability of receiving placebo, (*2*) low illness severity, and (*3*) more recent RCTs to be associated with greater placebo response. Papakostas et al. reviewed 23 relevant meta-analyses (of which 12 original, not yet included in the two previous meta-reviews), and reported repetitive evidence for a positive association between placebo response and (*1*) lower probability of receiving placebo, (*2*) low illness severity, and (*3*) increased visit frequency. The authors of three meta-analyses (at study-level), published between 2004 and 2010, unanimously concluded that at least a lower probability of receiving placebo is likely to inflate placebo response in depression trials (Khan, Kolts, Thase, Krishnan, & Brown, 2004; Papakostas & Fava, 2009; Sinyor et al., 2010). However, the results of four more recent meta-analyses (of which three at study-level) published between 2012 and 2016, strongly indicate that there is no such effect (Dunlop et al., 2012; Furukawa et al., 2016; Iovieno & Papakostas, 2012; Mancini, Wade, Perugi, Lenox-Smith, & Schacht, 2014). The inconsistency in results shows that, even when repeated findings lead to seemingly justifiable conclusions, subsequent meta-analyses exploring the same relationship may generate conflicting results, especially when data are aggregated at study-level. It underlines the need for authors of reviews to collect data from as many sources as possible, and to preferably weigh the results of individual studies on the basis of certain quality criteria.

1.1. Aims of the study

The objective of the current work is to review the collective evidence regarding associations of placebo response with trial design and patient characteristics that have been previously explored in pooled analyses, meta-regressions, and other effect models whereby the results of RCTs in MDD are combined. Unlike previous systematic reviews of meta-analyses, the aim of the current paper is to bring together the results of all previous work and grade the strength of evidence for identified predictors on the basis of pre-defined criteria for quality, quantity, and specificity of the data underlying each analysis. This approach has several advantages. Firstly, the importance of contradictory or isolated findings, whenever occurring, can be weighed to a certain extent, which reduces the risk of drawing false or unjustified conclusions. It may also help to create a better understanding of the relatively weak predictive value of results coming from pooled analyses when these are based on aggregated data at study-level rather than at patient-level. This is important particularly when only RCTs are analyzed with relatively low placebo response (i.e., excluding negative, or failed studies which were never published), or when heterogeneity in the design of underlying RCTs is not well accounted for. Finally, it may contribute to a better trial design for the demonstration of efficacy of new products in depression.

2. Methods

The MEDLINE/PubMed publication database and the Cochrane Library were searched for meta-analyses and pooled-analyses (from here, all called 'meta-analyses' for the sake of simplicity) of placebo response in MDD. The search term 'placebo' was cross-referenced with the terms 'depression' or 'antidepressant,' 'response' or 'effect,' and 'trial' in Title/Abstracts to identify articles focusing on contributing factors to the placebo response, published in English between January 1990 and December 2017. Results were filtered to only show meta-analyses, reviews, and systematic reviews. Relevant abstracts were hand-searched, full articles obtained, and information from these utilized to synthesize the present systematic review. Reference lists of articles were also examined to identify further relevant studies not identified by the keyword searches. Meta-analyses that aimed to evaluate the association of study features with placebo response or the differential response to antidepressants and placebo were included in the current review, provided they were based on 'statistical aggregation' of (patient-level) data or (study-level) results from placebo-`controlled RCTs in depression. To be included, underlying RCTs were required to have enrolled patients with depressive symptoms, fulfilling further diagnostic criteria of MDD according to the Diagnostic and Statistical Manual of Mental Disorders (DSM, version III, III-R, IV, or IV-TR) or Research Diagnostic Criteria (RDC), and assessed with commonly accepted primary outcome variables such as the Hamilton Rating Scale for Depression (HAMD, 17 or 21-item version), Montgomery & Åsberg Depression Rating Scale (MADRS), and/or Clinical Global Impression scale (CGI, severity and/or improvement).(American Psychiatric Association, 1987, 1980, 1994, 2000; Guy, 1976; Hamilton, 1960; Montgomery & Asberg, 1979; Spitzer, Endicott, & Robins, 1978)

In order to evaluate the predictive strength of study outcomes, we assessed whether the meta-analyses (*1*) were based on a representative sample of RCTs, (*2*) focused on *illness severity* (improvement, or mean change in symptoms), or *curative effect* (percentage of participants fulfilling criteria for 'response' or 'remission') on placebo, rather than *trial outcome* (i.e., drug superiority over placebo, expressed in percentage of positive trials, or standardized mean difference between treatments) as endpoint, (*3*) applied formal and appropriate statistical testing, and (*4*) adhered to basic quality principles for meta-analysis. These four assessments are further explained below.

Ad (1). As far as could be verified, most of the meta-analyses were based on a sample of RCTs from two large and partially overlapping data sets that were included in two reference studies (Furukawa et al., 2016; Khan, Bhat, Kolts, Thase, & Brown, 2010). The two reference papers analyzed a total of 314 RCTs, testing the antidepressant qualities of 49 different drug formulations against placebo between the years 1978 and 2015. For each meta-analysis, the amount of underlying RCTs already listed in the two reference papers was used to calculate the Jaccard index (*T*) as a measure of overlap or representativeness, using the formula:

$$T = N_c/(N_a + N_b - N_c)$$

whereby N_a is the total number of underlying RCTs included in the meta-analysis, N_b is the total number of RCTs listed in the two reference papers ($N_b = 314$), and N_c is the number of RCTs in the meta-analysis that were also included in the two reference papers. When authors of a paper did not provide further details on RCTs underlying their meta-analysis, the maximum Jaccard index was calculated, assuming that all of the underlying RCTs already were included in the list of reference trials. In addition to the Jaccard index, the total number of trial participants exposed to placebo or active drug were extracted and tabulated, as well as the period in which underlying RCTs were completed or reported (whichever was mentioned).

Ad (2). For those meta-analyses in which a positive trial outcome or effect size (the difference between active drug and placebo) was used as an endpoint (rather than cure, or illness severity changes on placebo), results were considered to not provide direct evidence for an effect on placebo response.

Ad (3). Associations between explored variables and placebo response were only considered to be positive or negative when statistically significant under the reported testing conditions.

Ad (4). To further assess the scientific rigor of the meta-analyses, similar criteria were applied as on the basis of which the Overview Quality Assessment Questionnaire (OQAQ) was earlier validated as an index of the quality of review articles (Oxman & Guyatt, 1991). More specifically, this concerned the following criteria: (a) unpublished studies are included or searched for (comprehensive search); (b) search terms are clearly specified (selection bias was avoided); (c) descriptive data are presented for each study, such as design, subject