



Placebos and the Placebo Effect in Drug Trials

Paul Enck and Sibylle Klosterhalfen

Contents

1	Introduction	400
2	Placebo Effects and Placebo Efficacy in Drug Trials	401
2.1	Patient Contributions Towards the Placebo Effect	401
2.2	Doctor/Therapist Contributions Towards the Placebo Effect	406
2.3	The Contribution of Disease Characteristics	408
2.4	The Role of the Trial Designs and Characteristics	411
3	Traditional Concepts to Minimize Placebo Effects	413
3.1	Multiple Centers, Transnational	413
3.2	Placebo Run-Ins and Withdrawals	415
3.3	Enrichment Designs and Adaptive Designs	416
4	The Challenge of Omitting Placebos	418
4.1	Comparative Effectiveness Research (CER)	418
4.2	Waiting List Controls, Treatment as Usual, and Preference Designs	419
4.3	Open-Label (“Real-Life”) Observational Studies and Registry and Cohort Studies ...	419
5	Other Challenges for Future Studies	423
5.1	E-Health and m-Health	423
5.2	Placebo Effects with Personalized Medicines	424
6	Summary	425
	References	425

Abstract

In this review, we explored different ways of controlling the placebo effects in clinical trials and described various factors that may increase/decrease the placebo effect in randomized placebo-controlled trials. These factors can be subdivided into four groups, and while not all factors are effective in every study and under all clinical conditions, they show on the whole that – even

P. Enck (✉) · S. Klosterhalfen

Department of Internal Medicine VI: Psychosomatic Medicine and Psychotherapy,

University Hospital Tübingen, Tübingen, Germany

e-mail: paul.enck@uni-tuebingen.de

under the ideal condition of drug therapy, where blinded placebo provision is much easier and warranted than in, e.g., psychotherapy – many factors need to be controlled to ascertain that the goal of the clinical trials, fair assessment of superiority of the drug over placebo in placebo-controlled trials and fair assessment of non-inferiority of the drug compared to another drug in comparator trials, is reached. Ignorance towards the placebo effect, which was common in the past, is no longer acceptable; instead, it should be the goal of all therapeutic trials to minimize the placebo effect in clinical trials, while utilizing and maximizing it in clinical routine.

Keywords

Clinical trials · Control conditions · Design · Drug effect · Nocebo effect · Placebo effect

1 Introduction

In both traditional and modern pharmacology, placebos are understood as tools, as research vehicles with which the true efficacy and mechanism of action of “real” drugs can be elucidated. Although this tool has been around for over a century (Jutte 2013), it did not earn its rightful place in pharmacology until very recently. We could have known better placebo research commenced as early as in the 1940s, when Henry K. Beecher (1904–1976) reasoned about the size and the mechanisms of the “placebo effect” in the first placebo-controlled clinical trials of his time (Beecher 1955) and Steward Wolf (1914–2005) promoted experimental placebo studies in his milestone paper “pharmacology of the placebo” (sic!) in the prestigious *Pharmacological Reviews* in 1959 (Wolf 1959). In 1980, the editor of *Handbook of Experimental Pharmacology*, Volume 55/I, states in the preface that “the only real psychoactive drug is the placebo: it acts directly on the psyche” (Stille 1980).

In this chapter, we will base our discussion on the content of the *Handbook of Experimental Pharmacology*, Volume No. 225 of 2014 (Benedetti et al. 2014): Although the last 5 years may have brought some new details to light about novel aspects and sophisticated features of the placebo effect and the placebo response, most of what we know today about it is summarized in this reader, as well as in a number of other collections and books published within the last 5 years (Benedetti 2014; Colloca 2018a, b; Enck et al. 2019). For those interested in single studies and papers concerning the term, we refer to the *Journal of Interdisciplinary Placebo Studies* (JIPS) literature database (www.jips.online) which, at present (2019), contains more than 4,000 genuine data papers and reviews on the placebo topic (Enck et al. 2018).

Limitations

Due to space limitations, this chapter will not discuss at length the history of the use of placebos in pharmacology (Kaptchuk 1998; Jutte 2013), nor will we refer in detail to the underlying mechanisms of the placebo effect/response, learning, and

expectations (Schedlowski et al. 2015). We will also refrain from exploring the neurophysiological and biological pathways involved in eliciting responses after placebo provision. Finally, we will abstain from discussing the placebo effects in non-drug therapies: physical therapy (Maddocks et al. 2016), psychotherapy (Enck et al. 2019), instrumental therapies (Burke et al. 2018), acupuncture (Chae et al. 2018), and surgery (Wartolowska et al. 2014) have their own specific and non-specific effects when tested against “sham” interventions, if these are feasible and acceptable. Furthermore, we do not intend to provide an answer to the question as to whether placebo pills (or equivalent medicinal preparations: drops, ointments, injections, infusions, enemas, etc.) are actually required to elicit the placebo response or whether verbal instructions alone are sufficient.

We will instead focus on issues relevant to drug development and drug testing and discuss the ways in which drug efficacy has been dealt with in clinical pharmacology in the past and present, how they may be handled in the future using placebos, and potential alternatives to its utilization. We will continue to bear in mind that the use of placebos has been questioned not only for ethical reasons. Finally, we will explore design alternatives that may be used for both experimental and clinical studies “in the real world” of the future. Albeit this constitutes an exploration of the ways in which placebo effects have affected drug testing, and not the changes of clinical trials in general during the last 25 years (May 2019), and reading through this summary will also identify many features that we discuss in the following chapter.

2 Placebo Effects and Placebo Efficacy in Drug Trials

Below, we will discuss four major factors that determine the placebo effects in drug trials: contributions from patients, contributions from doctors, the role of the disease and its characteristics, and, finally, the role of study designs and trial features. Before doing so, we like to emphasize that, whenever possible, we will distinguish placebo effects from spontaneous variation of symptoms but are aware of the fact that in both drug and placebo arms of randomized, placebo-controlled trials (RCT), the contribution of symptom variation is not always easy and sometimes impossible unless a “no-treatment” arm is included – which is hampered by ethical restraints and psychological barriers discussed later. Our basic understanding is illustrated in Fig. 1.

2.1 Patient Contributions Towards the Placebo Effect

2.1.1 Age and Sex

Among the earliest speculations that placebo effects in RCT are controlled by patient characteristics is the assumption that placebo effects are higher in younger patients

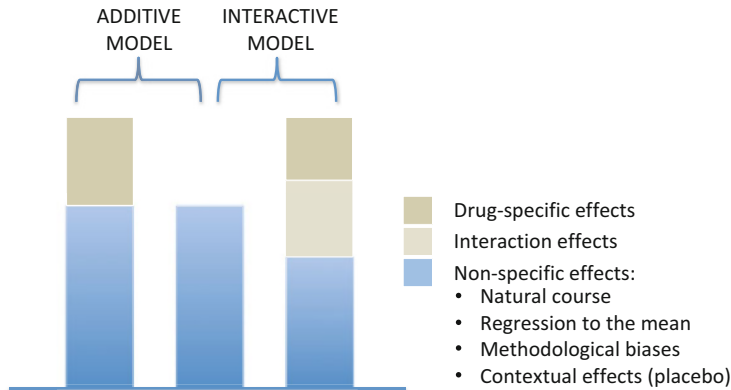


Fig. 1 The “additive model” in pharmacotherapy is the basis for all current drug therapy and its development: it assumes that by double-blinded randomization of patient to either the drug or the placebo arm of the trial, all other factors (natural course, regression to the mean, biases) are kept equally balanced between the two, and the same holds true for the contextual (placebo) effects. While this may be true in a global sense (Kirsch 2000), it has been questioned (Enck et al. 2011a, b), and evidence has been accumulated that at least in some cases, the biology of the placebo effect, e.g., release of endogenous endorphins in case of placebo analgesia, may interfere with the drug effect, e.g., of exogenous pain killers, and may either increase or decrease the placebo effect, leading to false estimation of the efficacy. This is illustrated with the “interactive model” (Enck et al. 2013a)

than in adults and in the elderly and that women show higher placebo responses than men. Both of these assumptions are, however, false.

In a systematic review of 75 meta-analyses on RCT across medicine (neurology, psychiatry, internal medicine) (Weimer et al. 2015a, b), we found only 20 in which an age effect of the placebo response was noted. In 15 analyses the response was said to be higher in younger patients, while in 5 the opposite effect was noted. This poor supportive evidence for an age effect is mainly derived from studies in children and adolescents (Weimer et al. 2013), while there are considerably more studies in adults. However, this effect may be due to specific modalities of pediatric RCT, while age effects among adults have rarely been shown. We have proposed a model (Fig. 2) that allows different developments depending on the type of disease but assumes that the overall response may be a stable pattern (type 2) once patients reach adulthood.

The situation is somewhat different with respect to gender: Again, our systematic review (Weimer et al. 2015a, b) did not support the notion that women show higher placebo effects than men, since only 3 of the 75 meta-analyses noted any gender differences at all. However, evidence from experimental placebo research, either specifically addressing the sex issue or accidentally finding sex differences, left us with a different impression: According to one systematic review using placebo (pain/analgesia) models with verbal placebo instructions (Vambheim and Flaten 2017), the summary of the results of 18 experimental approaches showed evidence of a higher placebo response in males than in females, while the females reacted

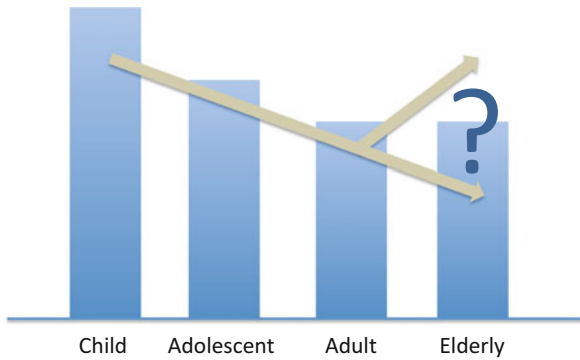


Fig. 2 The placebo effect with increasing age. Some data support that from childhood via adolescence to adulthood, the placebo effect decreases, at least in some clinical conditions (Weimer et al. 2013). We here speculate whether it further decreases at higher age due to decreased expectancy and relevance of the symptoms or whether it increases again with increased experience of effective therapy during the lifespan, based on a conditioning/learning hypothesis. Without further evidence, it is reasonable to assume that it stays stable at the level reached during adulthood

more strongly in conditioning (learning) experiments and with nocebo (symptom worsening) paradigms.

The apparent difference between experimental work on the one hand and clinical studies on the other hand lets us augment the systematic review (Enck and Klosterhalfen 2019) and hypothesize that this difference is due to the fact that, under laboratory conditions, the separation of learning (conditioning) mechanisms and verbal manipulation of expectancies is feasible and enables such differentiation. In clinical trials, however, patients are exposed to settings determining their expectations (e.g., informed consent about potential benefits and adverse effects of the treatment) but also bring their complete disease (or medicine, illness, treatment) history into this setting, thereby mixing learning and expectation mechanisms so that the net (placebo) effect does not permit the identification of the sex-specific relative contribution of each: There may be sex differences, but at the end of the day, these do not surface in RCT. And, as we will see below, this picture becomes even more distorted by the “placebo-by-proxy” effect.

2.1.2 Personality and Genes

Although there has already been much speculation over the years, the proof for a “placebo personality” (patients prone to respond to a placebo provision) remains rather weak (Kaptchuk et al. 2008). The reason is somewhat unexpected: Drug companies, when seeking approval for a novel drug in RCT during its development, do not tend to include psychometric tests to screen for personality profiles and/or specific psychometric characteristics – except in psychiatry and related areas, where psychiatric comorbidity may be part of the disease itself. This is because if the drug response depends at least partly upon psychometric scales, they are at risk of receiving a selective indication: No company would dare to do so. Furthermore, as

has been pointed out (Kaptchuk et al. 2008), to establish the existence of a behavioral response pattern “placebo responder,” the response needs to be shown to be stable across different trials and with different drugs for different diseases. Since this has rarely been tested clinically (Whalley et al. 2008) and has produced conflicting results (de la Fuente-Fernandez 2012), it thus disproves the concept. Even within a setting and a RCT, placebo run-in phases were unable to eliminate placebo responses during the trial (see below).

If anything, these data indicate that specific psychological traits are associated with higher (or lower) placebo response rates, coming as they do from experimental studies, albeit involving healthy volunteers. A number of characteristics that have been subject of systematic reviews are identified (Darragh et al. 2014; Horing et al. 2014). While several of these concepts, such as dispositional optimism (Geers et al. 2010), extraversion (Kelley et al. 2009), and an external locus of control (Horing et al. 2015), have even been replicated, it is a matter of some debate as to whether this renders them applicable to patient characteristics. It is, however, important to note that – contrary to common belief – higher placebo responses are associated with an “outward” orientation (externalization), while patients with high inward orientation (high self-efficacy) are less prone to respond to placebos.

In another study with a large group of healthy volunteers ($N = 624$) undergoing placebo analgesia/nocebo hyperalgesia induction by verbal suggestion plus experimental manipulation, a multivariate analysis of somatosensory and psychological variable reveals no predictive power for placebo responses, but personality traits such as neuroticism and extraversion as well as pain modulation by distraction and sex were able to predict nocebo hyperalgesia, the somatosensory response pattern being the strongest predictor of nocebo responses (Christian Büchel, Hamburg, personal communication).

Another reason for this poor outcome of psychometric screening for placebo responders may be of a methodological nature: The significant associations of single traits (or subscales of traits) reported may have been purely random and may be due to a beta error. Many tests were carried out, but only a few subscales – precisely those reported – yielded significance. A multivariate approach with a reasonably large sample may overcome such a bias.

While it is still too early for a final conclusion, the search for genes or polymorphisms of genes predicting the placebo response makes the same mistake: For whole-genome analyses (GWAS, genome-wide association studies), the samples are usually too small to allow adjustment for multiple comparisons, and candidate gene approaches replicate only what has been found for other psychological or behavioral traits and conditions. Summary reviews (Colagiuri et al. 2015; Hall et al. 2018) propose a “placebome” list an assembly of 28 genes/SNPs in 42 studies to date (Wang et al. 2017) to which more and more studies will be added in the future, albeit probably without improving the concept to any great extent.

2.1.3 Proxies

One of the most neglected research areas in placebo research, with far-reaching effects on placebo responses, is the influence of the social environment of the patient, relatives, and friends and, specifically, of other patients with the same or with other

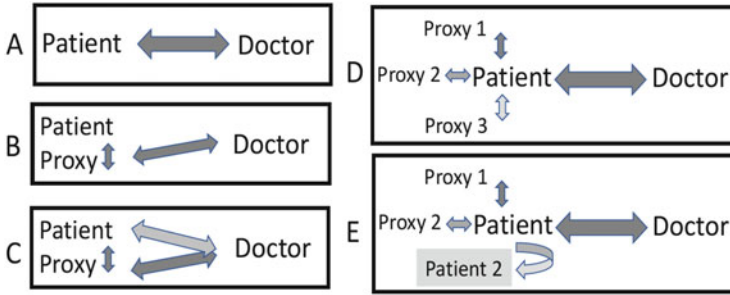


Fig. 3 The “placebo-by-proxy” concept (Grelotti and Kaptchuk 2011) illustrated in a systematic way, in which placebo responses are generated by increasing complexity of the network of interactions (different shades of gray reflect different communication intensities). (a) An idealized medical situation in contemporary medicine, where the (adult) patient individually communicates with the doctor and reports all relevant events in his/her medical history and environment (including family). Our understanding of the placebo effect is typically based on this constellation. (b) The concept illustrated reflects where the patient may experience limitations to direct communication with the doctor, due to verbal (infants, animals), social (migrant), or cognitive (intellectual disability) limitations. Proxy reports, based on either observation of the patient behavior or on (limited or special) communication strategies, are required. (c) Instead of exclusively communicating with the proxy, doctors may rely on additional information directly from the patient. This may generate conflicting information, e.g., higher placebo effects from proxy reports than from measures. (d) The social environment of a patient usually contains more than one proxy, with varying proximities to the patient, from family (parents, children, siblings) to relatives and friends/peers/colleagues. Proximity determines how much they may be involved in the medical history and its reporting and how much the doctor may be aware of this social network and its influence on disease reporting, management, and efficacy. (e) It is conceivable that one or more of the members of a social network may also have an impact as patient, though the timing and direction of effect may not be readily apparent but via an iterative process become contributors to the treatment effect of the index patient, either via social observation or explicit or implicit learning and vice versa

diseases. This concept has been called “placebo by proxy” (Grelotti and Kaptchuk 2011) and is observed when patients are unable to directly express their symptoms and symptom changes to their physician, instead of requiring a “proxy” to do so: these are predominantly children and mentally disabled.

We summarized this concept and developed a kind of systematic classification (Fig. 3) for future studies. For the time being, however, we are left with a few empirical examples demonstrating its clinical relevance. Our concept may also account for the differences observed between patient and proxy ratings of symptom improvement, e.g., in attention deficit hyperactivity disorder (ADHD) (Waschbusch et al. 2009).

One novel variant of the placebo-by-proxy concept will be discussed later, but the increasing use of social media and Internet fora by patients recruited for drug studies causes concern among trialists, e.g., with respect to the quality blinding in RCT (Lipset 2014); its impact on testing drug and placebo efficacy still needs to be determined.

2.2 Doctor/Therapist Contributions Towards the Placebo Effect

2.2.1 Age, Sex, and Ethnicity

Until the late 1980s, most RCTs in common diseases, where patient recruitment is not difficult to achieve, were monocentric, and thus the question as to what extent the placebo effects are attributable to the individual treating physician could not be answered: center effects on RCT outcome were simply not discernible and therefore of no consequence. This may be the real reason why everybody seems to believe that placebo responders may exist (Benedetti and Frisaldi 2014): Placebo producers, doctors who were able to push both placebo and drug effects up higher, were appreciated rather than dismissed and were rarely challenged.

However, even in individual centers, patients are often treated by different physicians. In a post hoc analysis of a RCT for treatment of irritable bowel syndrome (IBS) (Enck et al. 2005a, b), we had access to individualized patient and doctor data and ascertained that the female physician generated a better outcome than her two male colleagues in both the diet and drug and in the placebo arm of the study. Similar data resulted from an acupuncture trial in which female acupuncture therapists were more frequently believed to have administered true (as opposed to sham) acupuncture in a controlled acupuncture trial than their male counterparts (White et al. 2003): Female physicians appear to elicit more trust than their male counterparts.

While this phenomenon is well established in social psychology for most types of day-to-day communication among individuals, it had not yet been tested extensively in patient-doctor interaction. In experimental placebo research, female experimenters were observed to produce higher placebo analgesia rates (i.e., reports of less pain) in male volunteers, but not in females (Aslaksen et al. 2007); in experimental nausea and placebo/nocebo responses, we often noted sex-by-sex interactions of the outcome of respective studies (Enck and Klosterhalfen 2019), as already discussed above (Sect. 2.1.1) with regard to sex differences on placebo response in general. What is more, in a series of such nausea studies with German and Chinese volunteers (Klosterhalfen et al. 2005a, b, 2006), one female Chinese experimenter was unable to secure reliable nausea reports from her colleagues because they were (male) students, while she was a university teacher in China.

While systematic exploration of such factors in RCT is wanting, basic experiments pave the way: doctor ethnicity and gender affect patient judgment to a high degree, resulting in variable trust scores and the willingness to believe and comply (Shah and Ogden 2006). A simulation study comprised 300 UK patients who rated each one of 8 pictures of doctors of varying sex (male, female), age (young, old), and race (Asian, Caucasian) with respect to their anticipated personal manners, technical skill explanatory skills, advice, emotional aspects, and referral behavior, all of which are liable to contribute to placebo responses. They described remarkable differences – particularly between gender and race – with respect to patient expectation, but not necessarily with respect to true consulting behavior.

2.2.2 Training, Education, and Communication Skills

Little is known about how the medical training of doctors contributes to the response of patients during a RCT in general, let alone the specific response to placebo in a trial such as this. One ingenious experiment at Harvard Medical School (Jensen et al. 2014) sheds some indirect light on this question: Doctors in training were recruited for a brain imaging study in which they were told that the purpose of the study is to ascertain how the treatment of a patient effectively influences the doctor's brain. The rest is camouflage: an instructed patient-actor performed "pain relief" and "pain worsening" following button-pressing of the doctor inside the scanner that mimics successful or failed pain blockade via a sham device on the patient's arm; the doctor was able to observe the reflection of the facial response in a mirror. The perceived pain relief was directly linked to activation of the reward areas (e.g., area postrema) in the doctor's brain, and these, in turn, were the very same areas (the so-called pain matrix) that are known to mirror placebo analgesia in patients, as shown in different experiments (Legrain et al. 2011). On the basis of such data, training medical students in doctor-patient interaction may have a profound influence in future RCT.

Our final illustration of the relevance of expectations is derived from a study conducted in a Canadian hospital, in which more than 300 patients were asked to rate the empathy of the treating doctor (on a standardized scale) when attending a clinic for a common cold (Rakel et al. 2009). Patients who perceived their physician as empathic were shown to have significantly less severe symptoms, and, as even laboratory tests confirmed, the duration of their cold was almost a day shorter.

Finally, training during preparation of a RCT to better standardize patients' communication and information is required. Failure of drug trials (Kobak et al. 2007) is often associated with poor preparatory training of doctors prior to the study, inadequate conductance (e.g., recruitment and treatment in the hands of the same person), and biased evaluation of treatment outcome, particularly when based on subjective measures by the treating physician. However, standardized patient assessment by independent raters, video-recorded control, and combined doctor- and patient-reported outcomes are still not universal standards. This may well explain reported discrepancies in placebo response rates in RCTs (in depression treatment) between PRO and doctor ratings (Rief et al. 2009a).

2.2.3 Setting

In a quasi-experimental study (incidental rebuilding of a medical outpatient center), architecture, design, and service, as well as seasonal variations, were shown to have the ability to substantially improve the response to medical treatment (Rehn and Schuster 2017). This serves to illustrate that many more factors than the immediate circumstances on drug/placebo provision contribute to the overall treatment effect, of which only a few, such as those related to the empathy communication skills of the therapists, may be standardized through training, as discussed above. Such "incidental effects" (Grünbaum 1986) are difficult to control and require careful inspection of the site, time, and the staff conducting the RCT.

While we acknowledge that many of these influential factors may be averaged out by selecting many centers, each of which recruits only a small fraction of patients for the RCT, it cannot be ruled out that the known nationality-dependent effects of different placebo response rates in different regions of the world (EU versus USA) in multinational trials may be due to such effects. The time spent at the first consultation in primary care can vary substantially from country to country, even in Western countries (Irving et al. 2017).

2.3 The Contribution of Disease Characteristics

2.3.1 Disease Severity

Disease severity is one of the major driving forces for placebo effects in RCTs: Our analysis of the placebo responses in psychiatric (Weimer et al. 2015a, b) and other RCTs (Weimer et al. 2015a, b) across different clinical conditions showed that a lower disease severity in almost all meta-analyses was associated with higher placebo responses. Lower symptom severity is therefore one of the very few factors that predict the placebo effect in both adults and children.

To lend support to this statement as a more general rule, we deem it necessary to define “severity” on the basis of disease symptoms rather than of disease biomarkers: At the time of its first clinical diagnosis, a disease that initially has only very few symptoms, for example, juvenile diabetes, may not respond to placebo application at all but could well respond to metabolic interventions. On the other hand, diseases with a high symptomatic load, such as asthma, may respond stronger to placebo interventions following a drug intervention affecting forced expiratory volume (Wechsler et al. 2011). This underlines the importance of subjective measures in addition to biomarkers for many, if not for all, conditions.

2.3.2 Disease Duration

In agreement with a low disease severity at the disease onset, a short medical history and disease duration have been found to be associated with higher placebo responses in RCTs (Weimer et al. 2015a, b). Although this may well be the driving factor for higher placebo responses at younger age (see above), it has never actually been evaluated. In a meta-analysis of pediatric depression trials, the same holds true for children: the lower the severity, the higher the placebo response (Bridge et al. 2009).

At this point, drug development may run into a paradox, a kind of “trap,” when selecting only mildly affected patients for treatment of a putatively chronic condition as early as possible and before the disease exacerbates: such secondary prevention trials may be at overestimation of their efficacy. The same phenomenon may occur if – for economic or marketing reasons – patients recruited for RCTs during drug development do not represent the majority of patients in clinical routine and the drug proves to be disappointing after marketing approval, as was the case with the class of serotonergic antidepressants (Kirsch 2016).

2.3.3 Previous Treatments

It had been noted already some time ago (Rickels et al. 1966) that a preceding treatment of a disease may co-determine the success or failure of a subsequent treatment and that this applies not only to drug effects but also to placebo effects and in both directions: Treatment success may predict higher responses, and treatment failure may result in lower responses in the next trial (Colloca and Benedetti 2006). This is highly compatible with the concept that the placebo response is a conditioned response – albeit conditioning and expectancy cannot be as easily differentiated in medical treatment as in the laboratory for experimental placebo studies (Enck et al. 2008).

At the same time, for many clinical conditions, a shorter disease history and presumably a lower disease severity at least in case of chronic diseases are known to be associated with higher placebo response rates in RCT (see above). The immediate consequence of this is an apparent paradox: Testing novel drugs in patients with less severe symptoms may generate better drug responses but drives the placebo response higher, and so larger sample sizes are then required to yield significance in RCT.

This is of great relevance for drug testing in many respects: To begin with, novel drugs tested successfully in RCT often disappoint in the real world once they compete with drugs on the market and are tested on patients who have experienced both success and failure. At the same time, *The Emperor's New Clothes* (Kirsch 2014) fuels expectations and makes counter-evidence and contradictory experience likely. Finally, this calls for inclusion of the patients' medical history, especially their previous drug treatments, into the screening procedure for RCT, including their participation in earlier drug testing RCTs. However, since the latter is at conflict with both ethical and legal rules, we will discuss a potential solution at a later stage in this paper.

2.3.4 Adverse Event Rate of Drugs

Each and every placebo-controlled trial assumes perfect blinding of the study medication, which is feasible provided that the company producing the drug is also responsible for the production of the (undistinguishable) placebo. Under these circumstances, adverse events (AE), particularly when based on subjective patient reports, occur to a similar degree in the two treatment arms (Mahr et al. 2017), and their overall incidence may not differ as long as the symptoms occurring are of a general nature (Rheker et al. 2018). This situation may change when the drug induces highly specific AE and side effects, but provided all potential AE are listed in the patient information and consent form, even those AE have a good chance of being listed under placebo conditions. A meta-analysis comparing AE reporting between different antidepressants (tricyclics, serotonin reuptake inhibitors) confirmed that in both the drug and placebo arm of the trials, it is the assessment procedure rather than the drug itself that determines the amount of AE reported and that the difference between the two drugs is also reflected in AE rates of the placebo arms in these studies (Rief et al. 2009b). This indicates that the information provided about AE rather than the actual occurrence of AE is the driving force of such “nocebo effects” (Enck et al. 2013a, b) in RCTs. As already illustrated in

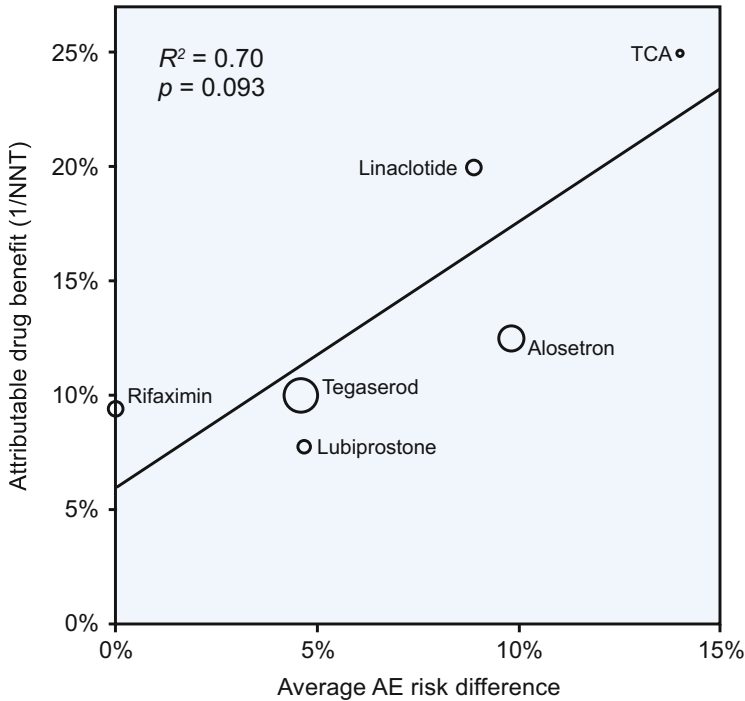


Fig. 4 Implicit unblinding of a study, based on reported adverse events (AE), as it becomes visible during a meta-analysis (Shah et al. 2014): Significant correlation between patient-reported efficacy and average adverse event risk difference in different drug therapies of irritable bowel syndrome (IBS). The size of each data point correlates with population size as a relative measure of variance for the assessment of adverse events. The positive and significant correlation indicates that with higher AE risk (difference between AE is the drug and the placebo arm of the RCT), the relative drug benefit (1/NNT) increases. (Reproduced with permission from Wiley and Sons, License No. 4627120830140)

meta-analyses, nocebo response rates determine the rates of discontinuation, e.g., in Parkinson’s disease (Leal Rato et al. 2019).

Such explicit unblinding – which may also occur when patients participating in a RCT communicate via social media (see below, Sect. 5.2) – is not to be confused with another phenomenon labeled “implicit unblinding” (Shah et al. 2014), which was identified in a meta-analysis of RCT in treatment of irritable bowel syndrome (IBS): the authors analyzed 6 different IBS treatment approaches in 30 RCT, either with (serotonergic) prokinetics (alosetron, linaclotide, tegaserod), with tricyclic antidepressants, sodium chloride channel blockers (lubiprostone), and with a locally acting antibiotic (rifaximin). In summary, they ascertained that the higher the reporting incidence of AE in the drug arm of these trials compared to placebo, the higher the reported drug-placebo difference (and, thus, the drug benefit). Figure 4 shows the correlation between the two, indicating implicit unblinding even if the individual patient in any of the trials is not aware of it.

2.4 The Role of the Trial Designs and Characteristics

2.4.1 Crossover Versus Parallel Group

In the early phases of drug development (the second half of the twentieth century), crossover trials were quite common – patients received either placebo or drug in a double-blinded manner in a first phase and then, following a washout period, the alternate application for the same duration. The advantage is each patient served as his/her “own” control, thus reducing data variance and enabling smaller numbers of patients to achieve statistical significance of drug over placebo. It also complied with an ethical stipulation that all patients should receive effective treatment, either immediately or after the placebo period.

The disadvantage is an effective treatment with the drug during the first phase affected the second treatment period – while the drug may have been washed out, conditioning effects are not, unless they are extinct (Suchman and Ader 1992). They increase the placebo effects over the “placebo-first” group. Similarly, if the drug was ineffective in the first phase, this had consequences for the placebo treatment that ensued. In consequence, treatment effects (drug and placebo, respectively) could be merged only if they were equipotent, irrespective of their order of provision; otherwise, only the first phase of treatment could be used for efficacy evaluation, and the advantage of the crossover would be lost, since it then would become a parallel-group designed study.

Nowadays, crossover designs are usually used to meet ethical requirements and to improve patient recruitment in cases in which leaving a patient with a placebo treatment only might be seen as unacceptable – for reasons medical, ethical, or psychological.

To wash out a conditioning effect in a crossover design study, it may be advisable to provide a placebo during the washout phase, as well as a kind of randomized withdrawal strategy, so that individual patients are switched from drug to placebo or vice versa, double-blinded, and with different timing (Moore et al. 2015) (Fig. 5). To the best of our knowledge, this has never been tested for feasibility in a crossover design study; it would still be necessary to control for equal starting out conditions in the two arms.

2.4.2 Trial Duration

In older textbooks of clinical pharmacology, you often will find the statement that placebo effects diminish all the more, the longer the trial lasts. In many RCTs in the last decade of the twentieth century, a conventional trial length lasted between 4 and 8 weeks, e.g., for acute conditions where a life-long intervention was not deemed necessary. In conditions prone to produce high placebo responses, e.g., in functional bowel disorders of IBS type (Elsenbruch and Enck 2015), it was proposed that trials lasting 8 weeks, which was common in the 1990s, should be extended to 12 weeks. The prediction was that this would result in lower placebo response rates in RCT (Spiller 1999) (Fig. 6).

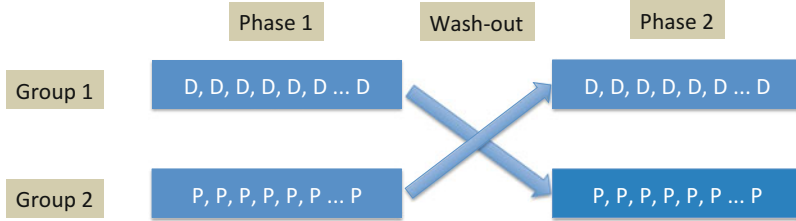


Fig. 5 A learning theory view on crossover trials with washout between drug and placebo phases. The unconditioned stimulus (US) is the drug (D), and the conditioning stimulus (CS) is the pill (shape, size, color, etc. = placebo). Groups 1 and 2 differ in the sequence they receive D and P; the washout phase may be of arbitrary length. In Group 1, the patient is conditioned in Phase 1 – by pairing the US and the CS – to respond to the CS alone in Phase 2: the washout period may eliminate the drug level, but it does not extinct the conditioned response unless a placebo (CS) is provided without the US. Thus, extinction will only gradually occur in Phase 2. In Group 2, the patient is initially exposed to the CS alone, a learning strategy which is called “latent inhibition” (Klosterhalfen et al. 2005a, b) that will minimize the conditioned response in Phase 2 – the washout phase does not serve any purpose. Therefore, while the two D phases may be comparable, the two P phases are not, and the calculation of the global drug efficacy based on intraindividual D-P differences is not an adequate estimation of it

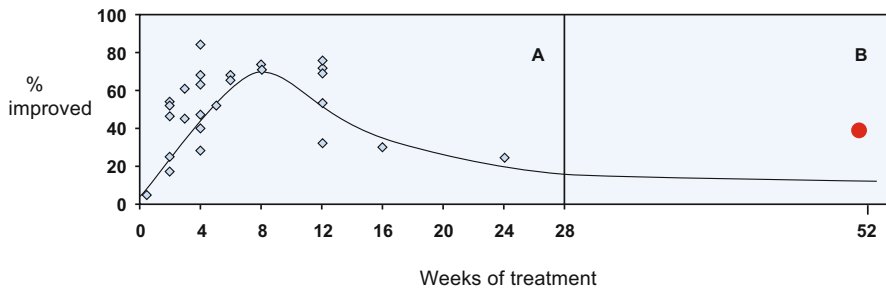


Fig. 6 Based on 26 randomized, placebo-controlled trials in irritable bowel syndrome (IBS) available at that time, it was argued (Spiller 1999) that with trial length over 24 week, placebo effects should reach a low level of 20% after half a year and decrease further afterwards (a). The extension of the plot beyond 28 weeks (b) was added to include the first 1-year study (Chey et al. 2004; red dot) in IBS with a stable 40% placebo effects for 1 year. (a) (Reproduced with permission from Excerpta Medica Inc., License No. 4627130163661)

However, when the first 12-week and longer trials were implemented, it became evident that placebo response could remain as high as 40% throughout such studies, and examples are available of 12-month trials with stable and high placebo response rates across the entire period (Khan et al. 2008; Quessy and Rowbotham 2008), not only in IBS (Chey et al. 2004).

The reason for this paradoxical prediction is that with a 4-week treatment trial, it may be possible to limit doctor-patient contacts to two – one at the beginning of the study and one at the end of trial – while for 12 weeks one would plan intermediate visits for motivation, compliance control, drug provision, and others.

By manipulating patients' expectancies; this increased number of contacts would reinforce the placebo effect (Enck et al. 2005a, b). And as with long-term, e.g., 1-year trials (Chey et al. 2004), the recording of symptoms and treatment effects would generally take the form of daily diary entries, phone calls from study nurses, and other measures. All these measures are liable to enhance the placebo effect, which is known to be driven by the extent of doctor-patient communication (Ford and Moayyedi 2010), irrespective of the nature of the disease (Jairath et al. 2016).

2.4.3 Randomization Ratio

If expectancy is another major driving force of the placebo effect in RCTs in addition to conditioning, the likelihood of receiving drug rather than placebo should affect the size of the placebo effect. A 50:50 randomization scheme is most common, but there are many reasons to deviate from it and to increase the percentage of patients in the drug arm of the study: for motivational reasons ("better than chance"), for ethical reasons (less patients without treatment), or to test different drug dosage in equally powered study arms against one placebo group.

It was first noted in a systematic review of migraine trials that increasing the chances of receiving active treatment causes the extent of the placebo effect to increase in a near-linear fashion (Diener et al. 1999). Subsequent analyses have confirmed this effect of "unbalanced randomization" in depression, in schizophrenia, and in other neurological and psychiatric conditions (Papakostas and Fava 2009; Mallinckrodt et al. 2010; Agid et al. 2013) (for a review see Weimer et al. 2015a, b). Interestingly, and for still unknown reasons, we were unable to confirm this phenomenon in the analysis of more than 100 RCTs in IBS (Elsenbruch and Enck 2015) (Fig. 7). Furthermore, unbalanced randomization does not influence the placebo effect in pediatric depression (Rutherford et al. 2011).

An easily conceivable endpoint of such study planning is reached when all (100%) patients receive active treatment and no placebo whatsoever is provided, such as with "comparative effectiveness research" (CER) or head-to-head trials, where a novel drug therapy is tested against another drug that is already available. We will discuss this further below, but a meta-analysis comparing efficacy of various antidepressants in placebo-controlled trials to CER studies using the same types of drugs revealed a 15% *higher* drug response in CER trials than in placebo-controlled trials (where the placebo response is on average 40%, Rutherford et al. 2009), which is solely attributable to the 100% anticipation of receiving active treatment.

3 Traditional Concepts to Minimize Placebo Effects

3.1 Multiple Centers, Transnational

Until the late 1990s, single-center studies were quite common in clinical drug testing, and there may still be a number of good reasons to maintain this tradition, e.g., in mechanistic studies in Phase II development or in the case of highly specific intervention strategies and modes, but definitely not for drug intervention. Center effects are thus avoided; they may be responsible for many drug failures once a drug

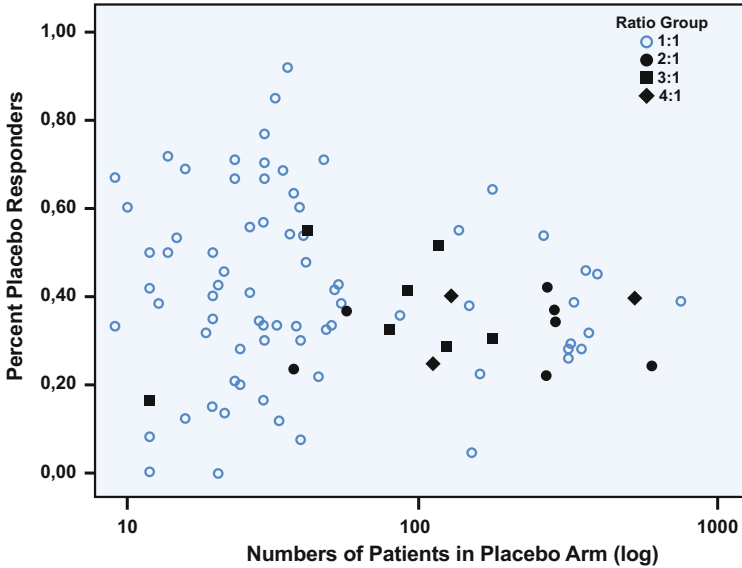


Fig. 7 The placebo effect in irritable bowel syndrome (IBS) trials as a function of the number of patients recruited: With higher patient numbers, the variance of the placebo effect between studies decreases and approximates 40% which has been found to be the global placebo response rates across all IBS trials (Ford and Moayyedi 2010). At the same time, the unbalanced randomization ratio (more patient assigned to drugs than to placebo) seems to not affect the placebo rates in IBS, while this has been found to be the case in depression, schizophrenia, and other conditions (Weimer and Enck 2014). (Reproduced with permission from Springer-Nature, License Number 4627150201527)

comes onto the market or even reaches the Phase III trials (Kobak 2010). Today's standards, multicenter trials with equal sample sizes and block randomization, may prevent overestimation of the drug-placebo difference to a considerable extent, albeit not completely: A higher number of study sites and a lower number of patients per study site were associated with higher placebo response (but not the drug response) in a meta-analysis of pediatric antidepressant trial (Bridge et al. 2009).

Extending multicenter trials across different countries is yet another option but one that bears many risks: Treatment of specific clinical conditions may be organized in very specific ways; hence, RCT results conducted in different countries could not be easily compared – and certainly not planned without taking the specifics of country, healthcare system, reimbursement policy, and alike into account. Cultural differences in the understanding (of the rationale for placebo-controlled trials) or interpretation (is it good to respond to placebo?) do exist (Ventriglio et al. 2018). Therefore, comparing placebo response rates – in meta-analyses – across different continents (Europe versus the USA) is crucial and has shown that overall European studies may generate higher placebo response, at least in some conditions (Stein et al. 2006). However, since neither Europe nor the USA is homogeneous cultural entity, subtle differences may sneak into individual RCT, depending on the range and location of recruitment centers.

3.2 Placebo Run-Ins and Withdrawals

The idea of ideally identifying putative placebo responders at an early point in a trial, or even before during recruitment of patients, is as logical as it is false: it assumes that being placebo responsive is a stable intraindividual characteristic that does not bear much empirical evidence (Kaptchuk et al. 2008). However, it bears another inherent risk: Being responsive to placebo does not rule out also being responsive to the drug, so by excluding responsive patients from the study, we may be preselecting the population, thereby introducing a selection bias; placebo responsiveness may thus indicate a subgroup of patients (such as those with lower symptom severity) and excluding these may put the requested indication for the drug at risk. A recent meta-analysis (Munkholm et al. 2019) indicates that placebo run-ins may also lead to false interpretation of drug efficacy: Participants treated with an antidepressant before recruitment and subsequently randomized to the study drug might experience withdrawal symptoms during the placebo run-in that are subsequently alleviated by the study drug.

We have already argued (above) that stable personality traits for placebo responsiveness do not exist. On the empirical-experimental side, the same person may be seen to respond to placebo provision in one trial, but not to another one in a different setting (Whalley et al. 2008). Furthermore, an effective treatment at one point in time may co-determine the response to any treatment (drug or placebo) on another occasion, both with experimental approaches (Colloca et al. 2010) and under clinical conditions (de la Fuente-Fernandez 2012), but is not warranted. The time frames for such “carry-over effects” have not been established, nor is it known how often a successful experience is required for it, how long it may last, and whether this also applies to negative (noneffective) treatment experiences (“nocebo”). The literature on Pavlovian learning is full of rules that may apply but that have yet to be explored.

In Fig. 5 (above), we have applied one such rule (extinction) to the test of carry-over effects in crossover trials. This resembles some similarities with randomized withdrawal studies, where patients are taken off the drug (or placebo) at the end of the trial in a blinded, randomized fashion (Fig. 8) to avoid conditioned rebound (nocebo) effects, i.e., effects that are due not to the pharmacologic withdrawal but to psychological effects such as disappointment at having reached the end of the study. This effect can be profound, as is shown in another example of the IBS literature (Chey et al. 2004): Having reaching the end of a 1-year study with persistent 40% placebo response and a stable 15% benefit above placebo in the respective arms, both drug and placebo recipients showed a dramatic recurrence of symptoms – a randomized withdrawal in the drug arm would presumably have shown a slower symptom worsening than in the placebo arm, which could be evaluated in terms of drug efficacy.

		Start of Trial	End of Trial
Drug Group	No. 1	P, D, D, D, D, D, D, D, D, D, P, P, P, P, P
	No. 2	P, P, D, D, D, D, D, D, D, D, D, P, P, P, P
	No. 3	P, P, P, D, D, D, D, D, D, D, D, D, P, P, P
	No. 4	P, P, P, P, D, D, D, D, D, D, D, D, D, P, P
	No. 5	P, P, P, P, P, D, D, D, D, D, D, D, D, D, P
		
Placebo Group	all pat.	P, P, P, P, P, P, P, P P, P, P, P, P, P, P, P

Fig. 8 The concept of randomized run-in and withdrawal in a clinical trial. To cover the true start and end of a trial, patients can be randomized to double-blinded run-in as well as withdrawal, where the true start of drug provision is hidden among days with placebo application instead. (Reproduced with permission from Springer-Nature, License Number 4627150201527)

3.3 Enrichment Designs and Adaptive Designs

Instead of removing putative or verified placebo responder, it was proposed that the group of drug responders be enriched during the course of study but without unblinding the study prematurely. The sequential parallel comparison design (SPCD) according to Fava et al. (2003) is quite an elegant attempt to overcome high placebo response rates, particularly in depression trials, for which it was originally developed. It operates in two phases, in which drug and placebo are unbalanced in favor of placebo, e.g., 1:2 or 1:3. Responders during this phase are removed to continue in an open fashion with whatever they had received. Non-responders are re-randomized to switch to the alternative (placebo and drug, respectively) and finish a second phase of the same length. At the end of the trial, data from both phases are pooled for statistical comparison in a conventional way for superiority of drug over placebo (Ivanova and Tamura 2015). This strategy (Silverman et al. 2018) allows better drug-placebo discrimination, even with placebo response rates as high as 40%, as is common in many clinical conditions with patient-reported outcomes (PRO). It is, to the best of our knowledge, the only patented design strategy that seeks to minimize placebo response and improve drug-placebo differences (assay sensitivity).

There are now many variants of the SPCD. A two-way enrichment design (Ivanova and Tamura 2015; Liu et al. 2019) re-randomizes drug responders and placebo non-responders during the first phase to a 50:50 drug: placebos are in a second phase – to maintain blinding until the very end – only the data from both the drug responders and the placebo non-responders from phase I are included in the analysis (Fig. 9). This is also thought to enrich the drug responders and can be combined with a randomized withdrawal strategy.

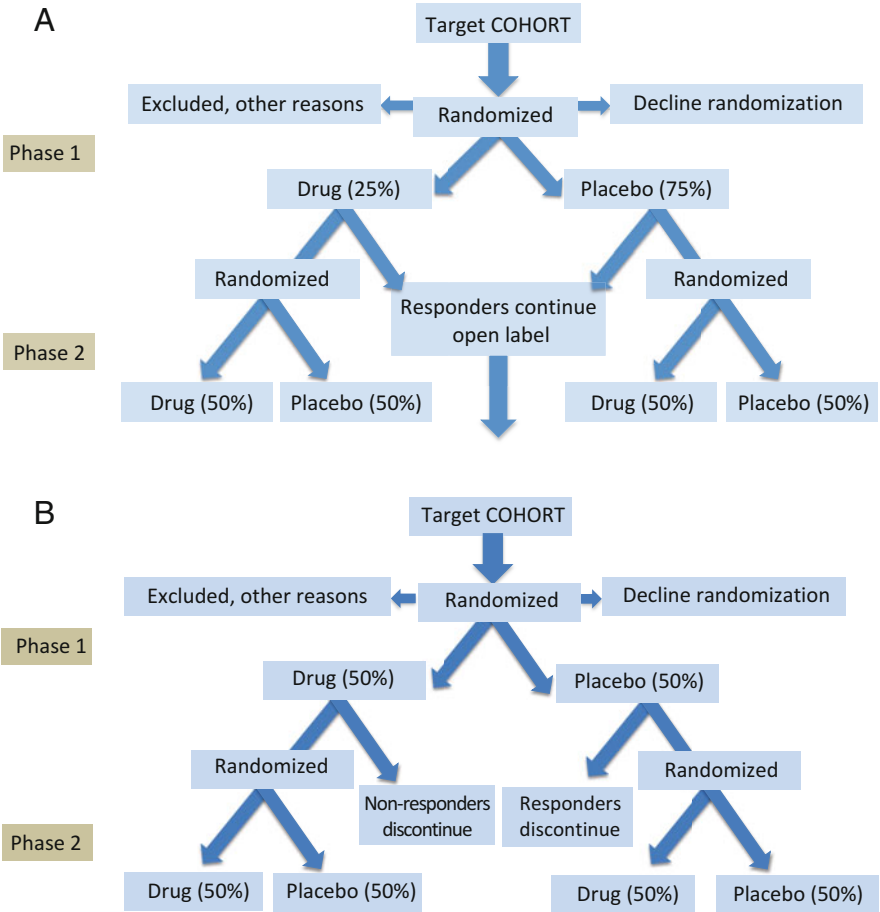


Fig. 9 Two enrichment designs to overcome increased placebo effects in RCT, especially in depression. (a) The original sequential parallel comparison design (SPCD) (Fava et al. 2003). The responders in both arms discontinue, and the non-responders are re-randomized to drug or placebo. Note that in Phase 1 more patients are randomized to placebo than to drug (2,1), while in the Phase 2 the randomization ratio is 1:1. (b) The two-way enrichment design (TED) (Ivanova and Tamura 2015) where the responders in the placebo arm and the non-responders in the drug arm are excluded while the respective others are re-randomized. Both strategies imply that – as long as both phases are equally long – the data of both phases can be merged to calculate the drug efficacy, but the results of Phase 1 are kept blinded until the very end

Other static or adaptive designs towards the same goal, such as the use of active placebos (Moncrieff et al. 2004; Jensen et al. 2017), have either been forgotten or are described in the literature but still await their clinical validation, e.g., the free-choice paradigm developed by our group (Enck et al. 2012), and our balanced crossover design (Enck et al. 2011a, b) eliminating limitations of conventional balanced

placebo design (Enck et al. 2013a, b). However, not all are suitable for validation in clinical trials, being predominately applicable predominantly in laboratory tests and trials.

4 The Challenge of Omitting Placebos

4.1 Comparative Effectiveness Research (CER)

Above (Sect. 2.4.3), we have already discussed the effects of increasing the likelihood of received active medication in placebo-controlled trials. While its extreme form – all patients receive active medication, either the drug under development or a comparator already on the market, thus having a 100% certainty of being treated by an active drug – may be favored by patients, ethics board, and approval authorities, it raises serious concerns among trialists: Omitting the placebo arm does not eliminate the placebo response but serves only to render it invisible and, therefore, uncontrollable. While we acknowledge its political and ethical intention, it is not without risk of seriously violating ethical and political rules at the same time. This is why:

- From a statistical standpoint, CER studies cannot hypothesize superiority of the novel compound over its comparator but can only claim (null hypothesis) non-inferiority (FDA 2016). However, non-inferiority requires an up to fourfold patient sample (for statistical reasons, see (Flight and Julious 2016)) and therefore violates the Declaration of Helsinki position that the least number of patients should be recruited for clinical trials, while all others should receive active medical care and treatment and not be exposed to medical research.
- CER studies require a comparator, but the choice among all *possible* comparators may co-determine the subsequent statistical testing and thereby the number of patients required to prove non-inferiority. Whether to select the best comparator on the market or an average comparative drug cannot, at the same time, be in the hands of the company developing the new drug nor in those of patients or patient representatives alone, as they may have divergent interests. It is therefore presumably an ethical issue to be decided by ethics boards or legal approval entities.
- Even if the requirement is to select the “best available comparator” on the market, this leaves a hole in the argument: should this be the best available drug on the market where the study is planned, or the best drug available globally, even if it is not available under these specific circumstances (country, healthcare system, clinic, or clinical condition), and who makes this decision? And what if the scientific community cannot even decide on account of different views on the evidence – should this again be decided by ethics board?
- And even if all these questions are answered: The drug under development needs to be indistinguishable from its comparator to allow a double-blinded assessment, and so both need to be produced by the same company, even if one is not its intellectual property. And who will force a company with a drug on the market that happens to be “the best comparator” to voluntarily provide its drug to a

competitor for such a testing that may turn out to be to its disadvantage? Is legal enforcement for such a policy required? Until these issues are solved, CER studies will not become pharmaceutical routine but are greatly dependent on a voluntary agreement among companies, ethics boards, and approval authorities. As was shown, most currently available (2019) non-inferiority trials are not appropriately designed to declare non-inferiority “even if it was worse than either placebo or another historic control” (Tsui et al. 2019).

4.2 Waiting List Controls, Treatment as Usual, and Preference Designs

One of the key issues of most, if not all, RCT designs is the fact that part of what occurs as placebo effect may be the consequence of spontaneous symptom variation and recovery – and it is generally assumed that the contribution of this factor to the overall effects (in both study arms) may be similar and can therefore be neglected when estimating the drug-placebo difference. This may also hold true in a similar way for CER studies.

However, with open-label observational studies, this becomes a factor of the utmost importance, since we are now dealing with one group only, and drug effects tend to be overestimated if non-specific contributions cannot be identified and enumerated. Conventional tools to overcome this limitation are waiting list controls and “treatment as usual,” but without proper randomization, they are subject to selection bias, either by the treating physician or by patients who have to agree to “treat or wait” or to novel versus conventional therapy. At the same time, symptom changes during waiting have been described in both directions (for the better, and for the worse) (Hesser et al. 2011; Furukawa et al. 2014). These were not the result of spontaneous symptom variation but were rather due to expectations and disappointment, respectively (Zhu et al. 2014). Waiting lists are generally used in psychotherapy where a blinded application of a sham intervention appears impossible (Gold et al. 2017) but are also used in some three-arm drug trials to control for spontaneous symptom variation (Krogsboll et al. 2009).

If treatment as usual and waiting list are used in RCT, however, they tend to reduce the non-specific effects due to disappointment and overestimate the efficacy of the therapy in the treatment arm (Fig. 10) (Enck and Lackner 2019). Rather than a single waiting list, a step-wedged waiting list (Fig. 11) may add value to this strategy by enabling to calculate a dose-response function for waiting. Patient motivation can be improved by preference designs when more than one type of treatment is available (Fig. 12), and the PD can also be applied to the CER strategy.

4.3 Open-Label (“Real-Life”) Observational Studies and Registry and Cohort Studies

Open-label observational studies were usually regarded as Phase IV marketing instruments of the drug industry, since their poor methodology provided little

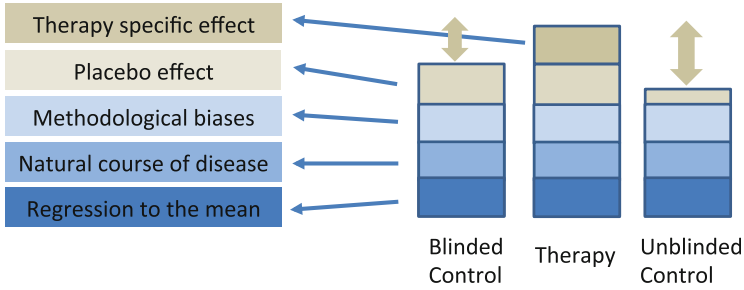


Fig. 10 The effect of unblinding in a RCT, such as with treatment as usual (TAU) and waiting list (WL) controls where blinding is impossible, e.g., in psychotherapy (Enck and Zipfel 2019), or where blinding is broken, e.g., due to AE reporting: The response in the control arm decreases and leads to overestimation of the efficacy in the treatment arm

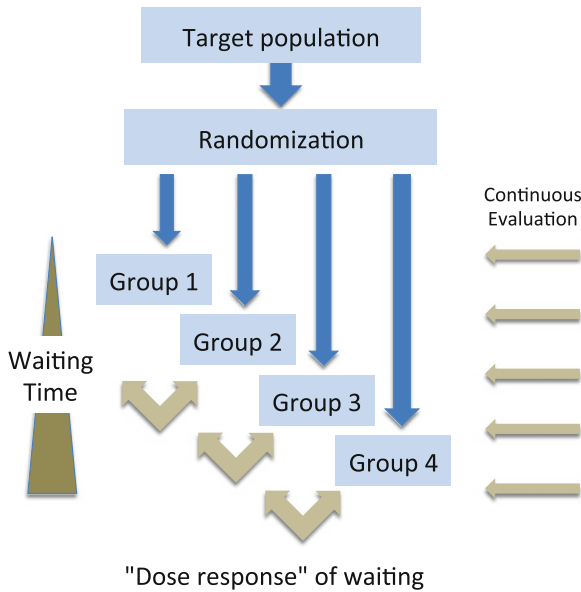


Fig. 11 A modified waiting list (WL) control strategy, where instead of one waiting list, two or more are implemented that reduce disappointment in patients randomized to WL (De Allegri et al. 2008) and allow the calculation of a waiting effect (as dose-response function) that can be separated from the placebo effect. (Reproduced with permission from Springer-Nature, License Number 4627150201527)

additional insight beyond what was known about drug efficacy at the time of approval and because they tended to substantially overestimate drug efficacy due to the lack of controlled conditions. This view changed once it became evident that patient selection during Phase III trials may also be biased – see our above arguments with respect to higher placebo response rates due to lower symptom severity in many

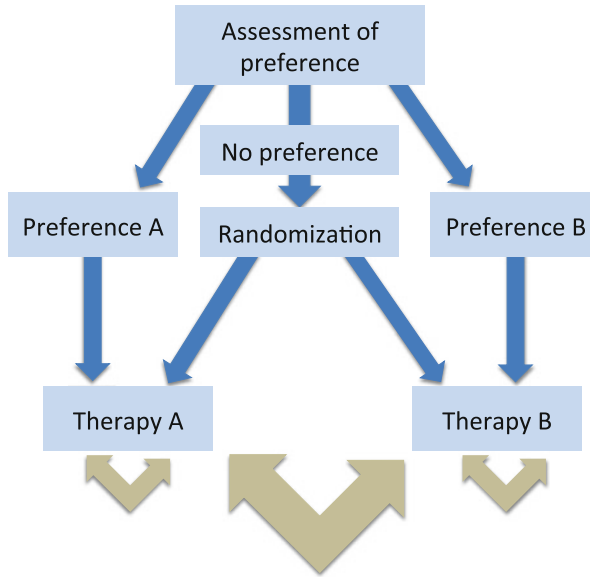


Fig. 12 A variant of a “preference design” where patients choose among true alternative treatments, and only those that do not report a preference are randomized to one of them. This can be applied to comparative effectiveness research (CER) studies where patient will receive a new drug or one already on the market or where true alternatives are to be compared, e.g., drug therapy versus surgery. It also allows comparison of efficacy between randomized and preference-assigned therapies. (Reproduced with permission from Springer-Nature, License Number 4627150201527)

clinical conditions. These may not represent those patients seen in private practices that do not participate in RCTs (the “real-world” patients) (Dal-Re et al. 2018).

In a bid to overcome these limitations, registry or cohort studies have been found helpful; at the same time, they make it possible to control spontaneous symptom variation in a very elegant way and without affecting patient motivation. An early design called “Zelen design” (Zelen 1979) was applied to all randomized placebo-controlled trials (Relton et al. 2010); here, we apply it to observational, Phase IV studies, to the best of our knowledge for the first time (Fig. 13).

Its basic idea is to recruit as many patients as possible for a “pure” observational study, either from a larger existing cohort or even a patient registry; the observational period needs to be defined and justified but can be of any length and recording frequency. The larger the cohort, the better it enables us to identify subgroups, e.g., with specific sociographic or clinical characteristics, specific treatment history, etc. These patients are asked to agree to a symptom monitoring for an extended period of time, but no interference with their ongoing therapy is envisioned (Phase I).

Once the recruitment is settled, the patients who have agreed to participate in the monitoring study are asked again whether they would consider volunteering for an interventional study (Phase II). This can be either a placebo-controlled trial or a CER trial. In both cases, the remaining observation-only group can serve as no-treatment

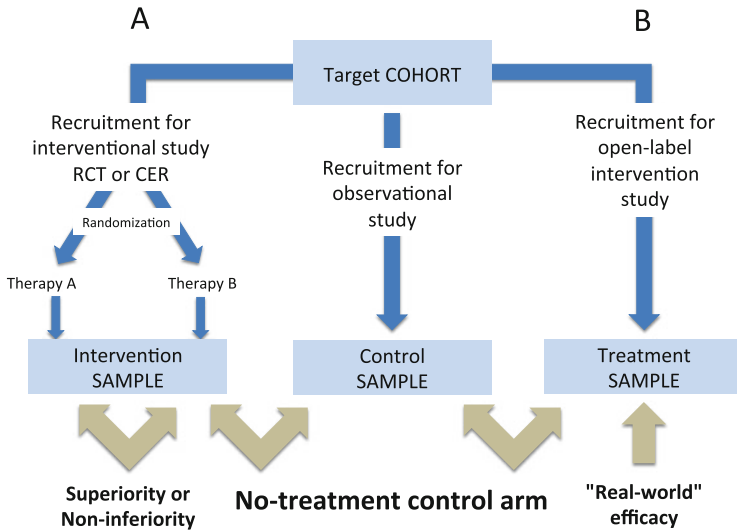


Fig. 13 A design alternative to address and calculate the effect of spontaneous symptom variation on drug and placebo effects which otherwise would require to randomize a patient to a “no-treatment” control, e.g., to a waiting list. The basic idea is to recruit a large number of patients to an “observation-only” study with fixed conditions (e.g., duration, number of observations, clinical conditions) (Zelen 1979; Relton et al. 2010), preferentially from a large patient cohort (registry). In a second step, those that have agreed to take part are subsequently asked whether they would participate in a conventional placebo-controlled or comparative study (a) or in an open-label study (b), and those not agreeing stay in the “observation-only” group. The larger the initially recruited cohort, the better a match between patients is feasible. Note that this allows a control group even in “real-world” studies, where otherwise controls are impossible to implement – we propose this “controlled open-label trial” (COLT) to overcome the limitations of pure observational studies

control and, if large enough, may even be matched to the treatment cohort with respect to sociographic or clinical criteria.

The “cohort multiple randomized controlled trial” (CMRCT) could even be applied to an observational study and would be the first of its kind to allow proper control of spontaneous symptom variation in observational studies without randomizing patients to a “no-treatment control”; we might call this the “controlled open-label trial” (COLT). Although this would at least give us some idea of the size of the “true” drug effect, we would still need to estimate the size of the contributing placebo effect, e.g., the difference between drug effect sizes in RCTs and in COLT-type studies.

5 Other Challenges for Future Studies

5.1 E-Health and m-Health

As discussed above, increasing the amount and intensity of study center (nurse, doctor) communication with patients is one of the driving factors of higher placebo response rates in some RCTs across medicine, with specific tools such as electronic symptom diaries, app-based reminders, random assessment of treatment effects, and chat rooms for patients to speak to their doctor or nurse when specific problems such as AEs arise.

At the same time, the vast amount of medical information available over the Internet (fact or faked) has dramatically changed the patient-doctor communication in daily practice and in RCTs: AE reporting is now highly correlated with the amount of websites discussing AE, e.g., of biosimilars versus biologics (Macaluso et al. 2018) and statins (Khan et al. 2018). This controls the (expectancy-mediated) “nocebo effect” of drugs and lowers patients’ willingness to participate in switch trials (Bakalos and Zintzaras 2018). The same holds true for the switch from branded to non-branded, generic products (Faasse et al. 2013).

More than a quarter of a million medical apps are currently available for various purposes. These include monitoring of treatment success/failure in placebo-controlled trials and in medical routine (FDA 2015), but a systematic evaluation of media-driven placebo effects is still lacking, even in laboratory settings and experiments. However, media-assisted provision e.g., of psychotherapy (by telephone, Internet, computer programs), can be as successful as face-to-face therapy, thus underlining that “digital placebo effects” (Torous and Firth 2016) are at least in a similar range, if not higher in those akin to these media – with more to come in the future: Just imagine having virtual doctors/nurses (Horing et al. 2016), patient avatars, and telemetric, wearable diagnostic and therapeutic tools.

The very same tools, specifically social media, interest groups, and chat rooms, have been found to be ideal if patients wish to exchange views with other patients recruited for the same study. Once this has been established, it may allow them to easily break any blinding code simply by accumulating AE and their frequency, provided that enough patients partake in the discussion. Unblinding, as we have shown (see above, Fig. 11), may not increase, but actually decrease the response to placebo, leading to overestimation of the drug effect as long as the source and size of unblinding remain undisclosed. Instead of fearing such development, doctors and researchers should take an active role to control such effects in the future.

Since patient recruitment has been professionalized over the past decade, social networks and media have taken over the recruitment of patients, e.g., by websites such as “Just Another Lab Rat!™” (www.jalr.org). This further supports what has been called “guinea-pigging”, uncontrolled participation of semi-professional volunteers as well as patients in more than one study at a time, or to overrule restrictions for further participation after completion of one study for the next 3, 6, or more months. Until there is a legal basis for a “study patient/volunteer registry” –

controlling recruitment and preventing its misuse but protecting both patient and drug company interests at the same time – the rapid technological development that we currently experience will leave traditional RCT methodology far behind.

5.2 Placebo Effects with Personalized Medicines

One of the promises of high-end medicine, or at least its current vision, is to provide personalized medicine, drugs developed for just one individual patient (or a subgroup of patients) whose genome has been used to develop and design the therapy. Whether this hails the end of the current mode of drug therapy testing is just one open question – with regard to the potential of placebo effects, it may certainly be seen as regressing back into the late nineteenth century: Individualization of therapy was, and still is, the premise of homeopathy and other complementary and alternative medicinal approaches, whether rational and justified or not (Mathie et al. 2018). We therefore expect, as in homeopathy, rising placebo response rates, at least for patient-reported outcomes; fortunately, most personalized therapies are developed initially for diseases with strong biomarkers (such as cancer) that are much less prone to placebo effects.

At the same time, personalized therapy – by definition – prevents the therapy from being controlled for placebo effects, e.g., against a standardized, nonindividual therapy (treatment as usual, for instance, or best therapy available). Even if groups of patients with common genomic markers, identified for a specific, personalized therapy, were to undergo such therapy, it is hard to think of a placebo or otherwise controlled condition that could be justified in terms of ethics, motivation, and costs. One way out of this dilemma would be the revitalization of $N = 1$ methodology (Kronish et al. 2018) that has developed its own strategies of proof-of-principle studies and statistical evaluation using, for example, time series analysis (Shaffer et al. 2018) to prove efficacy for one patient.

One completely different way of avoiding placebo controls was recently described by a drug company (Desai et al. 2013): they screened their entire archive of previously performed RCTs for studies where patients were recruited into a placebo arm of pain trials. After screening and merging the data (which had been stored in different databases) and screening for core data available in all studies, they were left with 203 studies with “historic” controls (called ePlacebo patients) treated with placebo. The idea is that these historic controls be used as a database rather than recruiting future patients into placebo arms of RCTs with novel compounds. The feasibility of such an approach, however, still needs to be verified prospectively, and has recently been questioned, as it may require substantially larger sample sizes as controls, especially with low effect sizes (Schoenfeld et al. 2019).

Table 1 A list of factors that have been found to be associated with the size of the placebo effect in RCT and meta-analyses

Patient characteristics	Doctor characteristics	Disease characteristics	Study characteristics
Age	Age	Baseline severity	Trial duration
Sex	Sex	Duration of illness	Number of sites
Personality	Race	Symptom load	Number of patients
Genes	Education	Previous therapy	Randomization ratio
Proxies	Empathy	Change during run-in	Number of visits
Education	Setting	Outcome measure	Type of assessment
Race	Behavior	Type of intervention	Industry support
Compliance			Type of control
Smoking status			Country/location
			Year of study

Note that the direction of change, e.g., higher placebo effects with higher or lower age, is ignored because of conflicting data and that some factors may have only been identified in only a few trials, e.g., industry sponsorship, while others, e.g., lower symptom severity at baseline, have been found in many RCTs to be associated with higher placebo effects (summary, based on Weimer et al. 2015a, b)

6 Summary

In this review, we have explored different ways of controlling the placebo effects in clinical trials and have described various factors that may increase/decrease the placebo effect in RCTs. As illustrated in Table 1, these factors can be subdivided into four groups, and while not all factors are effective in every study and under all clinical conditions, they show on the whole that – even under the ideal condition of drug therapy, where blinded placebo provision is much easier and warranted than in, e.g., psychotherapy (Enck et al. 2019) – many factors need to be controlled to ascertain that the goal of the clinical trials, fair assessment of superiority of the drug over placebo in RCTs and fair assessment of non-inferiority of the drug compared to another drug in CER trials, is reached. Ignorance towards the placebo effect, which was common in the past, is no longer acceptable; instead, it should be the goal of all therapeutic trials to minimize the placebo effect in clinical trials, while utilizing and maximizing it in clinical routine (Enck et al. 2013a).

Acknowledgments The assistance of Astrid Bahun from Bahun + Design for the artwork is gratefully acknowledged.

References

- Agid O, Siu CO, Potkin SG, Kapur S, Watsky E, Vanderburg D, Zipursky RB, Remington G (2013) Meta-regression analysis of placebo response in antipsychotic trials, 1970-2010. *Am J Psychiatry* 170(11):1335–1344

- Aslaksen PM, Myrbakk IN, Hoifodt RS, Flaten MA (2007) The effect of experimenter gender on autonomic and subjective responses to pain stimuli. *Pain* 129(3):260–268
- Bakalos G, Zintzaras E (2018) Drug discontinuation in studies including a switch from an originator to a biosimilar monoclonal antibody: a systematic literature review. *Clin Ther* 41(1):155–173
- Beecher HK (1955) The powerful placebo. *J Am Med Assoc* 159(17):1602–1606
- Benedetti F (2014) Placebo effects. Oxford University Press, Oxford
- Benedetti F, Frisaldi E (2014) Creating placebo responders and nonresponders in the laboratory: boons and banes. *Pain Manag* 4(3):165–167
- Benedetti F, Enck P, Frisaldi E, Schedlowski M (eds) (2014) Placebo, Handbook of experimental pharmacology, vol 225. Springer, Heidelberg
- Bridge JA, Birmaher B, Iyengar S, Barbe RP, Brent DA (2009) Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *Am J Psychiatry* 166(1):42–49
- Burke MJ, Kaptchuk TJ, Pascual-Leone A (2018) Challenges of differential placebo effects in contemporary medicine: the example of brain stimulation. *Ann Neurol* 85(1):12–20
- Chae Y, Lee YS, Enck P (2018) How placebo needles differ from placebo pills? *Front Psych* 9:243
- Chey WD, Chey WY, Heath AT, Dukes GE, Carter EG, Northcutt A, Ameen VZ (2004) Long-term safety and efficacy of alosetron in women with severe diarrhea-predominant irritable bowel syndrome. *Am J Gastroenterol* 99(11):2195–2203
- Colagiuri B, Schenk LA, Kessler MD, Dorsey SG, Colloca L (2015) The placebo effect: from concepts to genes. *Neuroscience* 307:171–190
- Colloca L (2018a) Preface (part I). In: *Neurobiology of the placebo effect part I*. Academic Press, Cambridge, pp xv–xx
- Colloca L (2018b) Preface (part II). In: *Neurobiology of the placebo effect part II*. Academic Press, Cambridge, pp xvii–xxiii
- Colloca L, Benedetti F (2006) How prior experience shapes placebo analgesia. *Pain* 124(1–2):126–133
- Colloca L, Petrovic P, Wager TD, Ingvar M, Benedetti F (2010) How the number of learning trials affects placebo and nocebo responses. *Pain* 151(2):430–439
- Dal-Re R, Janiaud P, Ioannidis JPA (2018) Real-world evidence: how pragmatic are randomized controlled trials labeled as pragmatic? *BMC Med* 16(1):49
- Darragh M, Booth RJ, Considine NS (2014) Investigating the ‘placebo personality’ outside the pain paradigm. *J Psychosom Res* 76(5):414–421
- De Allegri M, Pokhrel S, Becher H, Dong H, Mansmann U, Kouyate B, Kynast-Wolf G, Gbangou A, Sanon M, Bridges J, Sauerborn R (2008) Step-wedge cluster-randomised community-based trials: an application to the study of the impact of community health insurance. *Health Res Policy Syst* 6:10
- de la Fuente-Fernandez R (2012) The powerful pre-treatment effect: placebo responses in restless legs syndrome trials. *Eur J Neurol* 19(10):1305–1310
- Desai JR, Bowen EA, Danielson MM, Allam RR, Cantor MN (2013) Creation and implementation of a historical controls database from randomized clinical trials. *J Am Med Inform Assoc* 20(e1): e162–e168
- Diener HC, Dowson AJ, Ferrari M, Nappi G, Tfelt-Hansen P (1999) Unbalanced randomization influences placebo response: scientific versus ethical issues around the use of placebo in migraine trials. *Cephalgia* 19(8):699–700
- Elsenbruch S, Enck P (2015) Placebo effects and their determinants in gastrointestinal disorders. *Nat Rev Gastroenterol Hepatol* 12(8):472–485
- Enck P, Klosterhalfen S (2019) Does sex/gender play a role in placebo and nocebo effects? Conflicting evidence from clinical trials and experimental studies. *Front Neurosci* 13:160
- Enck P, Lackner JM (2019) Cognitive behavioural therapy for IBS: results or treatment as usual? *Nat Rev Gastroenterol Hepatol*. In press. <https://doi.org/10.1038/s41575-019-0174-2>
- Enck P, Zipfel S (2019) Placebo effects in psychotherapy: a framework. *Front Psych* 10:456
- Enck P, Klosterhalfen S, Kruis W (2005a) Determination of placebo effect in irritable bowel syndrome. *Dtsch Med Wochenschr* 130(34–35):1934–1937

- Enck P, Klosterhalfen S, Kruis W (2005b) Factors affecting therapeutic placebo response rates in patients with irritable bowel syndrome. *Nat Clin Pract Gastroenterol Hepatol* 2(8):354–355
- Enck P, Benedetti F, Schedlowski M (2008) New insights into the placebo and nocebo responses. *Neuron* 59(2):195–206
- Enck P, Klosterhalfen S, Weimer K, Horing B, Zipfel S (2011a) The placebo response in clinical trials: more questions than answers. *Philos Trans R Soc Lond Ser B Biol Sci* 366(1572):1889–1895
- Enck P, Klosterhalfen S, Zipfel S (2011b) Novel study designs to investigate the placebo response. *BMC Med Res Methodol* 11:90
- Enck P, Grundy D, Klosterhalfen S (2012) A novel placebo-controlled clinical study design without ethical concerns – the free choice paradigm. *Med Hypotheses* 79(6):880–882
- Enck P, Bingel U, Schedlowski M, Rief W (2013a) The placebo response in medicine: minimize, maximize or personalize? *Nat Rev Drug Discov* 12(3):191–204
- Enck P, Weimer K, Klosterhalfen S (2013b) Balanced placebo design, active placebos, and other design features for identifying, minimizing and characterizing the placebo response. In: Colloca L, Flaten MA, Meissner K (eds) *Placebo and pain: from bench to bedside*. Elsevier/Academic Press, Amsterdam, pp 159–173
- Enck P, Horing B, Broelz E, Weimer K (2018) Knowledge gaps in placebo research: with special reference to neurobiology. *Int Rev Neurobiol* 139:85–106
- Enck P, Weimer K, Colloca L, Doods S (eds) (2019) *Placebo and nocebo effects in psychiatry and beyond*. *Front Psychiatr*. Research Topic <https://www.frontiersin.org/research-topics/7727/placebo-and-nocebo-effects-in-psychiatry-and-beyond#articles>
- Faasse K, Cundy T, Gamble G, Petrie KJ (2013) The effect of an apparent change to a branded or generic medication on drug effectiveness and side effects. *Psychosom Med* 75(1):90–96
- Fava M, Evins AE, Dorer DJ, Schoenfeld DA (2003) The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom* 72(3):115–127
- FDA (2015) *Mobile medical applications – guidance for industry and food and drug administration staff*. Federal Drug Administration, Silver Spring
- FDA (2016) *Non-inferiority clinical trials to establish effectiveness – guidance for industry*. Federal Drug Administration, Silver Spring
- Flight L, Julious SA (2016) Practical guide to sample size calculations: non-inferiority and equivalence trials. *Pharm Stat* 15(1):80–89
- Ford AC, Moayyedi P (2010) Meta-analysis: factors affecting placebo response rate in the irritable bowel syndrome. *Aliment Pharmacol Ther* 32(2):144–158
- Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, Chen P, Hunot V, Churchill R (2014) Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr Scand* 130(3):181–192
- Geers AL, Wellman JA, Fowler SL, Helfer SG, France CR (2010) Dispositional optimism predicts placebo analgesia. *J Pain* 11(11):1165–1171
- Gold SM, Enck P, Hasselmann H, Friede T, Hegerl U, Mohr DC, Otte C (2017) Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. *Lancet Psychiatry* 4(9):725–732
- Grelotti DJ, Kaptchuk TJ (2011) Placebo by proxy. *BMJ* 343:d4345
- Grünbaum A (1986) The placebo concept in medicine and psychiatry. *Psychol Med* 16(1):19–38
- Hall KT, Loscalzo J, Kaptchuk T (2018) Pharmacogenomics and the placebo response. *ACS Chem Neurosci* 9(4):633–635
- Hesser H, Weise C, Rief W, Andersson G (2011) The effect of waiting: a meta-analysis of wait-list control groups in trials for tinnitus distress. *J Psychosom Res* 70(4):378–384
- Horing B, Weimer K, Muth ER, Enck P (2014) Prediction of placebo responses: a systematic review of the literature. *Front Psychol* 5:1079
- Horing B, Weimer K, Muth ER, Enck P (2015) Prediction of symptom change in placebo versus no-treatment group in experimentally induced motion sickness. *Appl Psychophysiol Biofeedback* 40(3):163–172

- Horing B, Newsome ND, Enck P, Babu SV, Muth ER (2016) A virtual experimenter to increase standardization for the investigation of placebo effects. *BMC Med Res Methodol* 16(1):84
- Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, Holden J (2017) International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ Open* 7(10):e017902
- Ivanova A, Tamura RN (2015) A two-way enriched clinical trial design: combining advantages of placebo lead-in and randomized withdrawal. *Stat Methods Med Res* 24(6):871–890
- Jairath V, Zou G, Parker CE, Macdonald JK, Mosli MH, Khanna R et al (2016) Systematic review and meta-analysis: placebo rates in induction and maintenance trials of ulcerative colitis. *J Crohns Colitis* 10(5):607–618
- Jensen KB, Petrovic P, Kerr CE, Kirsch I, Raicek J, Cheetham A, Spaeth R, Cook A, Gollub RL, Kong J, Kaptchuk TJ (2014) Sharing pain and relief: neural correlates of physicians during treatment of patients. *Mol Psychiatry* 19(3):392–398
- Jensen JS, Bielefeldt AO, Hrobjartsson A (2017) Active placebo control groups of pharmacological interventions were rarely used but merited serious consideration: a methodological overview. *J Clin Epidemiol* 87:35–46
- Jutte R (2013) The early history of the placebo. *Complement Ther Med* 21(2):94–97
- Kaptchuk TJ (1998) Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med* 72(3):389–433
- Kaptchuk TJ, Kelley JM, Deykin A, Wayne PM, Lasagna LC, Epstein IO, Kirsch I, Wechsler ME (2008) Do “placebo responders” exist? *Contemp Clin Trials* 29(4):587–595
- Kelley JM, Lembo AJ, Ablon JS, Villanueva JJ, Conboy LA, Levy R, Marci CD, Kerr CE, Kirsch I, Jacobson EE, Riess H, Kaptchuk TJ (2009) Patient and practitioner influences on the placebo effect in irritable bowel syndrome. *Psychosom Med* 71(7):789–797
- Khan A, Redding N, Brown WA (2008) The persistence of the placebo response in antidepressant clinical trials. *J Psychiatr Res* 42(10):791–796
- Khan S, Holbrook A, Shah BR (2018) Does Googling lead to statin intolerance? *Int J Cardiol* 262:25–27
- Kirsch I (2000) Are drug and placebo effects in depression additive? *Biol Psychiatry* 47(8):733–735
- Kirsch I (2014) The emperor’s new drugs: medication and placebo in the treatment of depression. *Handb Exp Pharmacol* 225:291–303
- Kirsch I (2016) The placebo effect in the treatment of depression. *Verhaltenstherapie* 26:55–61
- Klosterhalfen S, Kellermann S, Pan F, Stockhorst U, Hall G, Enck P (2005a) Effects of ethnicity and gender on motion sickness susceptibility. *Aviat Space Environ Med* 76(11):1051–1057
- Klosterhalfen S, Kellermann S, Stockhorst U, Wolf J, Kirschbaum C, Hall G, Enck P (2005b) Latent inhibition of rotation chair-induced nausea in healthy male and female volunteers. *Psychosom Med* 67(2):335–340
- Klosterhalfen S, Pan F, Kellermann S, Enck P (2006) Gender and race as determinants of nausea induced by circularvection. *Gend Med* 3(3):236–242
- Kobak KA (2010) Inaccuracy in clinical trials: effects and methods to control inaccuracy. *Curr Alzheimer Res* 7(7):637–641
- Kobak KA, Kane JM, Thase ME, Nierenberg AA (2007) Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol* 27(1):1–5
- Krogsboll LT, Hrobjartsson A, Gotzsche PC (2009) Spontaneous improvement in randomised clinical trials: meta-analysis of three-armed trials comparing no treatment, placebo and active intervention. *BMC Med Res Methodol* 9:1
- Kronish IM, Hampsey M, Falzon L, Konrad B, Davidson KW (2018) Personalized (N-of-1) trials for depression: a systematic review. *J Clin Psychopharmacol* 38(3):218–225
- Leal Rato M, Duarte GS, Ferreira AN, Alves M, Mainoli B, Teodoro T, Mestre TA, Costa J, Ferreira JJ (2019) Nocebo response in Parkinson’s disease: a systematic review and meta-analysis. *Parkinsonism Relat Disord*. In press. <https://doi.org/10.1016/j.parkreldis.2019.04.015>

- Legrain V, Iannetti GD, Plaghki L, Mouraux A (2011) The pain matrix reloaded: a salience detection system for the body. *Prog Neurobiol* 93(1):111–124
- Lipset CH (2014) Engage with research participants about social media. *Nat Med* 20(3):231
- Liu Y, Rybin D, Heeren TC, Doros G (2019) Comparison of novel methods in two-way enriched clinical trial design. *Stat Med*. In press. <https://doi.org/10.1002/sim.8288>
- Macaluso FS, Maida M, Ventimiglia M, Renna S, Cottone M, Orlando A (2018) Factors affecting clinical and endoscopic outcomes of placebo arm in trials of biologics and small molecule drugs in ulcerative colitis: a meta-analysis. *Inflamm Bowel Dis* 25(6):987–997
- Maddocks M, Kerry R, Turner A, Howick J (2016) Problematic placebos in physical therapy trials. *J Eval Clin Pract* 22(4):598–602
- Mahr A, Golmard C, Pham E, Lordache L, Deville L, Faure P (2017) Types, frequencies, and burden of nonspecific adverse events of drugs: analysis of randomized placebo-controlled clinical trials. *Pharmacoepidemiol Drug Saf* 26(7):731–741
- Mallinckrodt CH, Zhang L, Prucka WR, Millen BA (2010) Signal detection and placebo response in schizophrenia: parallels with depression. *Psychopharmacol Bull* 43(1):53–72
- Mathie RT, Ulbrich-Zurni S, Viksveen P, Roberts ER, Baitson ES, Legg LA, Davidson JRT (2018) Systematic review and meta-analysis of randomised, other-than-placebo controlled, trials of individualised homeopathic treatment. *Homeopathy* 107(4):229–243
- May M (2019) Twenty-five ways clinical trials have changed in the last 25 years. *Nat Med* 25(1):2–5
- Moncrieff J, Wessely S, Hardy R (2004) Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev* 1:Cd003012
- Moore RA, Wiffen PJ, Eccleston C, Derry S, Baron R, Bell RF, Furlan AD, Gilron I, Haroutounian S, Katz NP, Lipman AG, Morley S, Peloso PM, Quessy SN, Seers K, Strassels SA, Straube S (2015) Systematic review of enriched enrolment, randomised withdrawal trial designs in chronic pain: a new framework for design and reporting. *Pain* 156(8):1382–1395
- Munkholm K, Paludan-Muller AS, Boesen K (2019) Considering the methodological limitations in the evidence base of antidepressants for depression: a reanalysis of a network meta-analysis. *BMJ Open* 9(6):e024886
- Papakostas GI, Fava M (2009) Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol* 19(1):34–40
- Quessy SN, Rowbotham MC (2008) Placebo response in neuropathic pain trials. *Pain* 138(3):479–483
- Rakel DP, Hoelt TJ, Barrett BP, Cheung BA, Craig BM, Niu M (2009) Practitioner empathy and the duration of the common cold. *Fam Med* 41(7):494–501
- Rehn J, Schuster K (2017) Clinic design as placebo-using design to promote healing and support treatments. *Behav Sci* 7(4):E77
- Relton C, Torgerson D, O’Cathain A, Nicholl J (2010) Rethinking pragmatic randomised controlled trials: introducing the “cohort multiple randomised controlled trial” design. *BMJ* 340:c1066
- Rheker J, Rief W, Doering BK, Winkler A (2018) Assessment of adverse events in clinical drug trials: identifying amitriptyline’s placebo- and baseline-controlled side effects. *Exp Clin Psychopharmacol* 26(3):320–326
- Rickels K, Lipman R, Raab E (1966) Previous medication, duration of illness and placebo response. *J Nerv Ment Dis* 142(6):548–554
- Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG (2009a) Meta-analysis of the placebo response in antidepressant trials. *J Affect Disord* 118(1–3):1–8
- Rief W, Nestoriuc Y, von Lilienfeld-Toal A, Dogan I, Schreiber F, Hofmann SG, Barsky AJ, Avorn J (2009b) Differences in adverse effect reporting in placebo groups in SSRI and tricyclic antidepressant trials: a systematic review and meta-analysis. *Drug Saf* 32(11):1041–1056
- Rutherford BR, Sneed JR, Roose SP (2009) Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychother Psychosom* 78(3):172–181

- Rutherford BR, Sneed JR, Tandler JM, Rindskopf D, Peterson BS, Roose SP (2011) Deconstructing pediatric depression trials: an analysis of the effects of expectancy and therapeutic contact. *J Am Acad Child Adolesc Psychiatry* 50(8):782–795
- Schedlowski M, Enck P, Rief W, Bingel U (2015) Neuro-bio-behavioral mechanisms of placebo and Nocebo responses: implications for clinical trials and clinical practice. *Pharmacol Rev* 67(3):697–730
- Schoenfeld DA, Finkelstein DM, Macklin E, Zach N, Ennst DL, Taylor AA, Atassi N (2019) Design and analysis of a clinical trial using previous trials as historical control. *Clin Trials*. <https://doi.org/10.1177/1740774519858914>
- Shaffer JA, Kronish IM, Falzon L, Cheung YK, Davidson KW (2018) N-of-1 randomized intervention trials in Health Psychology: a systematic review and methodology critique. *Ann Behav Med* 52(9):731–742
- Shah R, Ogden J (2006) ‘What’s in a face?’ The role of doctor ethnicity, age and gender in the formation of patients’ judgements: an experimental study. *Patient Educ Couns* 60(2):136–141
- Shah E, Triantafyllou K, Hana AA, Pimentel M (2014) Adverse events appear to unblind clinical trials in irritable bowel syndrome. *Neurogastroenterol Motil* 26(4):482–488
- Silverman RK, Ivanova A, Fine J (2018) Sequential parallel comparison design with binary and time-to-event outcomes. *Stat Med* 37(9):1454–1466
- Spiller RC (1999) Problems and challenges in the design of irritable bowel syndrome clinical trials: experience from published trials. *Am J Med* 107(5a):91s–97s
- Stein DJ, Baldwin DS, Dolberg OT, Despiegel N, Bandelow B (2006) Which factors predict placebo response in anxiety disorders and major depression? An analysis of placebo-controlled studies of escitalopram. *J Clin Psychiatry* 67(11):1741–1746
- Stille G (1980) Preface. In: *Handbook of experimental pharmacology*, vol 55(I). Springer, Berlin, p V
- Suchman AL, Ader R (1992) Classic conditioning and placebo effects in crossover studies. *Clin Pharmacol Ther* 52(4):372–377
- Torous J, Firth J (2016) The digital placebo effect: mobile mental health meets clinical psychiatry. *Lancet Psychiatry* 3(2):100–102
- Tsui M, Rehal S, Jairath V, Kahan BC (2019) Most non-inferiority trials were not designed to preserve active comparator treatment effects. *J Clin Epidemiol* 110:82–89
- Vambheim SM, Flaten MA (2017) A systematic review of sex differences in the placebo and the nocebo effect. *J Pain Res* 10:1831–1839
- Ventriglio A, Magnifico G, Borraccino L, Rinaldi A, Bellomo A (2018) Placebo and cultural responses. *Nord J Psychiatry* 72(supp 1):S33–S35
- Wang RS, Hall KT, Giulianini F, Passow D, Kaptchuk TJ, Loscalzo J (2017) Network analysis of the genomic basis of the placebo effect. *JCI Insight* 2(11):93911
- Wartolowska K, Judge A, Hopewell S, Collins GS, Dean BJ, Rombach I, Brindley D, Savulescu J, Beard DJ, Carr AJ (2014) Use of placebo controls in the evaluation of surgery: systematic review. *BMJ* 348:g3253
- Waschbusch DA, Pelham WE Jr, Waxmonsky J, Johnston C (2009) Are there placebo effects in the medication treatment of children with attention-deficit hyperactivity disorder? *J Dev Behav Pediatr* 30(2):158–168
- Wechsler ME, Kelley JM, Boyd IO, Dutile S, Marigowda G, Kirsch I, Israel E, Kaptchuk TJ (2011) Active albuterol or placebo, sham acupuncture, or no intervention in asthma. *N Engl J Med* 365(2):119–126
- Weimer K, Enck P (2014) Traditional and innovative experimental and clinical trial designs and their advantages and pitfalls. *Handb Exp Pharmacol* 225:237–272
- Weimer K, Gulewitsch MD, Schlarb AA, Schwille-Kiuntke J, Klosterhalfen S, Enck P (2013) Placebo effects in children: a review. *Pediatr Res* 74(1):96–102
- Weimer K, Colloca L, Enck P (2015a) Age and sex as moderators of the placebo response – an evaluation of systematic reviews and meta-analyses across medicine. *Gerontology* 61(2):97–108
- Weimer K, Colloca L, Enck P (2015b) Placebo effects in psychiatry: mediators and moderators. *Lancet Psychiatry* 2(3):246–257

- Whalley B, Hyland ME, Kirsch I (2008) Consistency of the placebo effect. *J Psychosom Res* 64(5):537–541
- White P, Lewith G, Hopwood V, Prescott P (2003) The placebo needle, is it a valid and convincing placebo for use in acupuncture trials? A randomised, single-blind, cross-over pilot trial. *Pain* 106(3):401–409
- Wolf S (1959) The pharmacology of placebos. *Pharmacol Rev* 11:689–704
- Zelen M (1979) A new design for randomized clinical trials. *N Engl J Med* 300(22):1242–1245
- Zhu Z, Zhang L, Jiang J, Li W, Cao X, Zhou Z, Zhang T, Li C (2014) Comparison of psychological placebo and waiting list control conditions in the assessment of cognitive behavioral therapy for the treatment of generalized anxiety disorder: a meta-analysis. *Shanghai Arch Psychiatry* 26(6):319–331