

---

# Traditional and Innovative Experimental and Clinical Trial Designs and Their Advantages and Pitfalls

Katja Weimer and Paul Enck

## Contents

1	Introduction .....	238
2	Experimental Study Designs to Explore the Placebo Response .....	240
2.1	Manipulating Timing .....	240
2.1.1	Open/Hidden Treatment Paradigm .....	240
2.1.2	Delayed Response Paradigm .....	242
2.2	Manipulating Information .....	243
2.2.1	The Balanced Placebo Design .....	243
2.2.2	The Balanced Crossover Design .....	244
2.2.3	Modifying the Chances to Receive Drug or Placebo .....	245
2.2.4	Inverse Enrichment .....	246
2.3	Habituation, Sensitization, Learning .....	247
2.3.1	“No-Treatment” Controls in the Laboratory .....	247
2.3.2	Providing Models (Social Learning) .....	248
2.3.3	Providing Reinforcement (Instrumental Learning) .....	249
2.4	Predicting Placebo Responders .....	249
2.5	Avoiding Ethical Conflicts .....	250
2.6	One Size Fits All? The “Free Choice” Paradigm .....	251
3	Clinical Designs to Explore the Placebo Effect .....	252
3.1	Identifying Placebo Responders .....	252
3.1.1	Crossover Designs .....	252
3.1.2	Placebo and Drug Run-Ins .....	253
3.1.3	Randomized Run-in/Withdrawal .....	254
3.1.4	Repetitive Drug Application Phases .....	255
3.2	Controlling the Natural Course of Disease .....	256

---

K. Weimer • P. Enck (✉)

Department of Psychosomatic Medicine, University Hospital Tübingen, Tübingen, Germany

e-mail: [paul.enck@uni-tuebingen.de](mailto:paul.enck@uni-tuebingen.de)

3.2.1	Waiting Lists, Treatment as Usual .....	256
3.2.2	The “Zelen Design” or the “Cohort Multiple Randomized Controlled Trial” .....	257
3.2.3	Registry Trials .....	258
3.3	Improving Assay Sensitivity .....	258
3.3.1	Effective Blinding .....	259
3.3.2	Active Placebos .....	259
3.4	Improving Trial Acceptability .....	260
3.4.1	Unbalanced Randomization .....	260
3.4.2	Step-Wedge Design .....	261
3.4.3	Preference Design .....	262
3.4.4	Cluster Randomization .....	263
3.5	Developing Individualized Medicine .....	264
3.5.1	Previous Drug History .....	264
3.5.2	“Placebo by Proxy” .....	264
3.5.3	Patient Registry .....	265
3.6	Dismissing Placebos in RCTs .....	266
3.6.1	Head-to-Head Trials and CER .....	266
3.6.2	Historic Controls .....	267
4	Summary .....	267
	References .....	267

## Abstract

Many study designs and design variants have been developed in the past to either overcome or enhance drug–placebo differences in clinical trials or to identify and characterize placebo responders in experimental studies. They share many commonalities as well as differences that are discussed here: the role of deception and ethical restrictions, habituation effects and the control of the natural course of disease, assay sensitivity testing and effective blinding, acceptability and motivation of patients and volunteers, and the development of individualized medicine. These are fostered by two opposite strategies: utilizing the beneficial aspects of the placebo response—and avoiding its negative counterpart, the nocebo effect—in medical routine for the benefit of patients, and minimizing—by controlling—the negative aspects of the placebo effect during drug development.

## Keywords

Trial designs • Experimental designs • Placebo effect • Control group

## 1 Introduction

The following chapter will present and discuss both traditional and innovative and novel approaches to study the placebo response and its underlying mechanisms in laboratory experiments and in the clinical setting, with healthy volunteers as well as with patients.

While it will discuss and exemplify the traditional randomized and placebo-controlled study design that is “gold standard” since the mid of the twentieth century, it will not go any further in history to elaborate on the origin of this concept—this has been done in the first chapter of this book.

It also will not elaborate on the ethical implications of the use of placebos in the laboratory and the clinics, as this is not the expertise of the authors and is described somewhere else in this book. However, some of the ethical implications of many of the old and as well as the new designs will be discussed where appropriate to demonstrate that new methodologies may be based on ethical grounds, but may also generate new ethical conflicts and dilemmas. Ethics is an implicit challenge in all research involving humans, healthy volunteers or patients and will never find a final solution, at least not in placebo research. Similarly, we will not discuss ethic-related aspects that refer to patient information and informed consent procedures for the same reason, and for paucity of data.

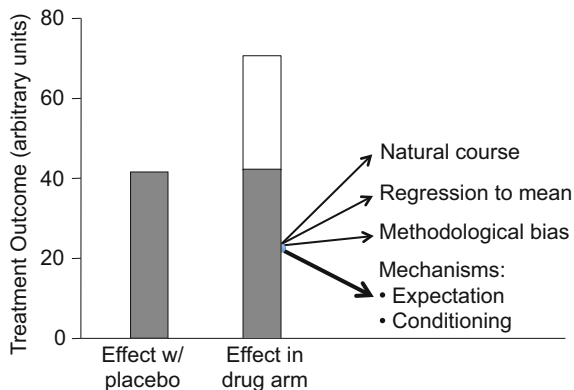
Dealing especially with experimental designs will bring this chapter close to the ones on mechanisms of the placebo responses, e.g., on learning and on expectations and conditioning, but we will not go into the details of it but restrict ourselves to issues where learning and expectations have specifically influenced design aspects. We have also excluded here studies where the purpose was dose reduction using conditioning paradigms via partial reinforcement as they are discussed elsewhere. We have finally excluded specifics of psychotherapy trials (except with respect to waiting list controls and their variants, see Sects. 3.2.1 and 3.4.2) because they represent a subset of study designs due to the fact that—different from all other nondrug interventions, e.g., surgery, physical therapy, and others—in psychotherapy the unspecific effects of drug therapy that include the placebo effects may become the specific effects of the psychological intervention (Kirsch 2005).

In the following, we will distinguish between experimental studies that are mostly performed in healthy volunteers but may also include patients, and clinical studies that are almost exclusively done in patients, at least once a drug is beyond Phase I of its development.

The latter studies are usually performed to compare a treatment (a drug, a nonpharmacological intervention, e.g., surgery) with a “sham” treatment (a placebo pill, sham-surgery, or other control procedures) to explore the benefit of the treatment above unspecific effects (often called placebo effects) that also include methodological biases, regression to the mean, and the spontaneous course of the disease (see Fig. 1).

The former are to explore mechanisms, and as such they may either explore mechanisms of action of the therapeutic intervention (drug, etc.) or of the placebo response. Only designs to explore the placebo response will be discussed here.

**Fig. 1** The “placebo effect” in both arms of RCTs is thought to be a composite of spontaneous disease variation, regression to the mean, and specific contextual factors that represent the placebo response. It is assumed that these factors contribute to both trial arms in an equal manner [adopted from Rief et al. (2011) with permission]



## 2 Experimental Study Designs to Explore the Placebo Response

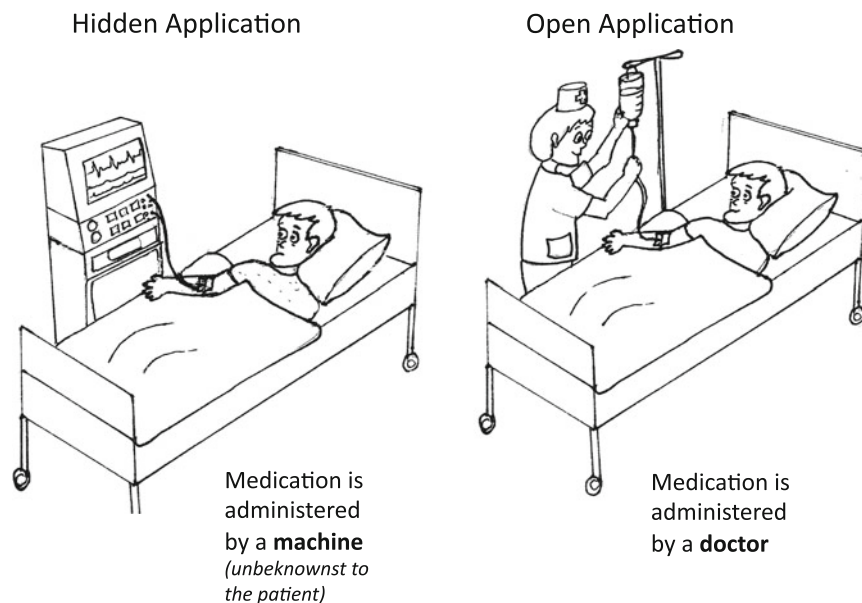
While in clinical studies the placebo effect is a compound effect of factors other than the placebo response of an individual (see Fig. 1), experimental designs in placebo research attempt to separate these components to—ideally—identify the “true” placebo effect. Two strategies can be singled out to do so: manipulating the timing of drug action and manipulating the information provided to the patients. The latter is much more common due to technical limitations of the first. Both carry specific ethical problems that will not be discussed here (see above).

### 2.1 Manipulating Timing

If placebo responses occur as an almost immediate consequence of a medical intervention intended to relieve symptoms in a patient as long as the patient expects symptom improvement to occur, placebo responses may even occur before a drug action can be noted. It has in fact been noted in experimental trials that the response in the placebo arm of a drug trial may be faster than in the drug arm in depression (Petrovic et al. 2002). Responses in short-term placebo or drug run-in phases in RCTs have been used to identify placebo or drug responders (see below Sect. 3.1.2). Therefore, dissociating the act of drug application from its presumed drug action onset in the eyes of patients allows separating the true (pharmacological) drug effect from the drug-plus-placebo effect in clinical trials. Two strategies can be found in the literature, of which only one has not yet found its way into experimental placebo research.

#### 2.1.1 Open/Hidden Treatment Paradigm

The open/hidden treatment paradigm (O/HP) was—based upon some empirical observations (Levine et al. 1978; Gracely et al. 1983)—developed by Benedetti and

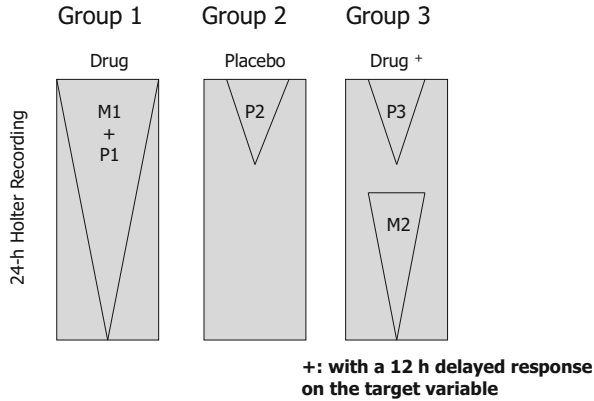


**Fig. 2** Open-hidden paradigm according to Benedetti et al. (2003): In this paradigm, identical concentrations of active drugs are administered by a physician in a visible (open condition) or hidden manner, in which the patient is unaware of the timing of administration of the medication (for example, a computer is used to control infusion timing). This permits the dissociation of the pure pharmacodynamic effect of the treatment (hidden treatment) from the additional benefit of the psychological context that comes from knowing that the treatment is being administered [adopted from Enck et al. (2013) with permission]

colleagues (Colloca et al. 2004; Benedetti et al. 2011) and demonstrates an exception from rules stated earlier: that studying the placebo response needs the application of a placebo. In the O/HP, no placebo is given but the timing of drug application is hidden to the patients allowing the placebo response to occur prior to the pharmacological action of the drug (Fig. 2). At the same time, this paradigm is presumably most effective with a real medical treatment situation, e.g., in treatment of acute pain.

Benedetti et al. have applied the paradigm in a number of clinical/experimental situations and have found that many drugs carry a substantial placebo effect in a standard medical setting where the open application of a drug is the rule eliciting strong patient expectations, including opioid and nonopioid analgesics (Amanzio et al. 2001), tranquilizers (Benedetti et al. 2003), and for a nonpharmacological intervention such as deep-brain stimulation in Parkinson's Disease (Pollo et al. 2002). The paradigm has also been used in experimental settings with healthy subjects undergoing pain simulation during brain imaging (Bingel et al. 2011).

While the O/HP may not be a suitable treatment model for clinical routine situations because it discourages the use of drugs with poor or questionable pharmacology, it carries a strong message into the clinics: even poorly effective



**Fig. 3** The “delayed response” design; M1 and M2 stand for medication response, P1 and P2 for placebo response; the “additive model” by Kirsch (2000) assumes that  $P1 = P2$ . Under the further assumptions that  $M1 = M2$  und  $P2 = P3$ , the hypothesis of the “additive model” is falsified if  $(M1 + P1 \neq M2 + P3)$  [adopted from Enck et al. (2011a) with permission]

drugs can show enhanced clinical efficacy when their open application makes use of the placebo response.

### 2.1.2 Delayed Response Paradigm

In the O/HP, the manipulation of timing is achieved via a computer-driven drug pump that randomizes (within given limits) the medicine application. In a theoretical model, we came to a similar—though presumably less reliable—technical solution by manipulating the drug release via tablet coating technology. It would dissociate the act of medication intake (swallowing a pill) from its pharmacological action and also allow the placebo response to occur prior to the true drug response; this was called the delayed response paradigm (DRP) (Enck et al. 2011a).

Different from the O/HP, the DRP would be most suitable specifically for drug studies in healthy participants and patients, both under experimental and clinical conditions, provided the pill coating technology would allow such procedures. However, it would require more than just one treatment group; ideally it would include 3 groups (Fig. 3) to identify the true drug, the true placebo response, and to verify the “additive model” (Kirsch 2000). All participants are informed that they will receive either a drug or placebo in a double-blinded fashion. No information, however, is provided about the timing of drug response but a cover story for the potential of prolonged drug action, e.g. for 24 h.

A variation of such a design that intended to elucidate the drug response in a clinical trial in Parkinson’s Disease was recently described (D’Agostino 2009): Patients in the placebo-arm are planned to switch from placebo to drug at some time point during the trial unbeknown to the patient and physician, but in this case pretreatment with placebo may affect the later drug treatment by conditioning procedures (Suchman and Ader 1992). A better way of separating drug and placebo effects may be randomized run-in and withdrawal periods (see below, Sect. 3.1.2).

## 2.2 Manipulating Information

Manipulation of information provided to volunteers and patients appears easier and is therefore most frequently done in placebo research—however, deception is evident in these cases and requires careful ethical consideration and approval, while with manipulation of drug timing (above), even fully informed consent may be possible.

In the majority of all experimental studies of the placebo response, the experimental group (that receives placebo) is usually provided with a 100 % security that the applied drug (pill, cream, injection, infusion, etc.) contains an effective pharmacological agent, while in fact they receive a placebo. In contrast, in clinical RCTs, patients usually receive the information that they have a 50 % (or another) chance to receive the active compound. The difference between both types of information accounts for substantially (up to sixfold) higher placebo effect size in the laboratory compared to a RCT (Vase et al. 2002), thereby allowing a better study of the underlying mechanisms. The control group serves as “no-treatment control” (see below, Sects. 2.3.1 and 3.2) and does not receive any treatment.

The downside of this common practice is the fact that the investigator is usually not blinded towards group assignment and treatment, and thereby may allow the response to be biased by implicit information and behaviors. Strictly separating data collection and data evaluation, or even using uninformed experimenters may help avoiding such bias but are not easy to establish. In the following we will present four experimental approaches to overcome these limitations.

### 2.2.1 The Balanced Placebo Design

The “balanced placebo design” (BPD) was traditionally used in the testing for expectancy effects of frequently consumed everyday-drugs such as caffeine, nicotine, and alcohol (Kelemen and Kaighobadi 2007), more recently also with drugs such as cocaine (Volkow et al. 2003) and marijuana (Metrik et al. 2009).

While one-half of the study sample receives placebo and the other half the drug, half of each group receives correct information while the other half receives false information on the nature of their study condition (drug or placebo) immediately prior to drug testing, thus allowing to differentiate between the “true” drug effect (those receiving the drug but are told they received placebo) and the “true” placebo effect (those receiving placebo but are told they received the drug) (Fig. 4).

The central concept of the design is—similar to the O/HP—to separate the “true” effects of drug from expectancy effects that occur when participants and patients are given a pill with the information that it may or may not contain the active compound.

A recent paper (Lund et al. 2014) used the BPD explicitly to evaluate whether the assumption of additivity that is implicitly underlying all RCT (Kirsch 2000) is correct. They found that the sum of the “true” drug effect and the “true” placebo effect is larger than the conventional “drug plus placebo” effect in trials, allowing estimating that RCTs tend to underestimate the drug effect and falsifying the additivity hypothesis.

		Information	
		Medication	Placebo
Application	Medication	1: true positive	2: false negative
	Placebo	3: false positive	4: true negative

**Fig. 4** The “balanced placebo design” (BPD): All participants are told they participate in a double-blind parallel-group design study. After drug intake and immediately before testing half of the participants in each group are given false and correct information on what they received [adopted from Enck et al. (2011a) with permission]

A variant of the BPD is the “half BPD” in which all participants are given placebos, but half of them receive information that they receive the drug—this is a more common design in current placebo research, as it does not require approval for performing a drug study where the ethical and legal stakes are usually higher. However, effective double-blinding of such a study is difficult unless—as in a recent test in our laboratory (Weimer et al. 2013b)—the participants and the experimenter(s) conducting the study are made to believe that they participate in a fully BPD.

One of the pitfalls of the BPD is the fact that all participants are informed (either correctly or falsely) prior to testing whether and what they have received. In sceptical participants (especially in medical students), this may raise doubts about the truth of the information provided and may require additional measures, such as a reliable explanation why the information is given at all. This is usually done by informing them that once the drug is active, the information whether and what they received may no longer be relevant—however, the participants’ acceptance of such information is difficult to prove prior to the test, and its testing afterwards may be subject to other biases.

### 2.2.2 The Balanced Crossover Design

In an attempt to overcome the serious limitations of the BPD, we designed another strategy that may account for some of the BDP limitations (Enck et al. 2011a). Participants are divided into four groups, and all are told they participate in a conventional trial, in which they will receive both the drug and the placebo at two different occasions in a randomized and double-blinded crossover fashion. This was called the balanced crossover design (BCD).

However, only Groups 2 and 3 will be exposed to drug and placebo in a balanced way, that is half the participants will receive the drug first and the placebo at the second occasion, while the other half will receive first placebo and then the drug. Group 1 will receive the drug twice, and Group 4 will receive placebo twice instead (Fig. 5). In this case, Groups 2 and 3 represent the conventional trial design for drug and placebo effects.



		First medication application	
		drug	placebo
Second medication application	drug	1: drug - drug	2: placebo - drug
	placebo	3: drug - placebo	4: placebo placebo

**Fig. 5** The “balanced cross-over design” (BCD): All participants are told they participate in a double-blind crossover design study and will receive both drug and placebo; this is true for groups 2 and 3, while in groups 1 and 4 they receive twice the drug and the placebo, respectively [adopted from Enck et al. (2011a) with permission]

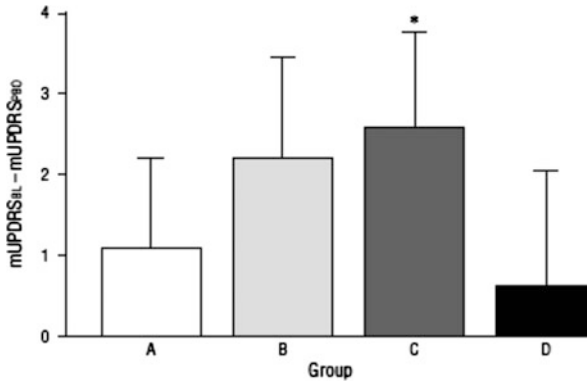
In Group 1, the minimal value of both measures represents the “true” drug effect (plus other unspecific effects), and the difference between both is the expectancy component of the drug response. In Group 4, the maximum value should represent the “true” placebo effect (plus other unspecific effects); and the difference between both values should be the expectancy component of the placebo response. Comparing these expectancy effects between groups 1 and 4 allows to test whether the expectancy component (the placebo effect) is equal under drug and placebo condition—which is the assumption of the “additive model”. All other nonspecific factors are assumed to be equally effective in all groups.

The balanced crossover design (BCD) has one important methodological limitation: As with other crossover designs, interference of learning effects need to be kept in mind (Suchman and Ader 1992; Colloca and Benedetti 2006; Kessner et al. 2013), and any adaptation or habituation between measurement 1 and measurement 2 should be minimized, e.g. by increasing the time interval between the two. Its ethical limitations (deception) are similar to those of the BPD with the exception, that participants may receive a drug twice but expect it to receive only once—any risk involved in such a repetition of drug application would exclude the BCD from use, and it can only be used in patients when the deception is authorized (Miller et al. 2005).

A study in our laboratory testing the effects of a nicotine patch on cognitive performance such as reaction times and response inhibition in healthy smoking and nonsmoking volunteers (Weimer et al. 2013c) showed its applicability and limitations.

### 2.2.3 Modifying the Chances to Receive Drug or Placebo

It has been shown that the likelihood of receiving the active treatment determines the size of both the drug and the placebo response in RCT (Papakostas and Fava 2009): the higher the likelihood of active treatment, the higher the response to both the drug as well as the placebo, solely attributable to the increased expectancy (Rutherford et al. 2009) (see below, Sect. 3.4.1). Maximal response difference between drug and placebo is achieved with a 50 % chance when the chances to receive either drug or placebo are equalized. This is thought to be associated with



**Fig. 6** Clinical response to placebo (modified Unified Parkinson Disease Rating Scale score at baseline [mUPDRS<sub>BL</sub>] – mUPDRS score following placebo [mUPDRS<sub>PBO</sub>]), adjusted for mUPDRS baseline and age. Values are given as mean (SD). There was no significant main effect of group. Only the change in group C was significant. \* $p < 0.05$ . In group A, subjects were told that their chances of receiving active levodopa were 25 %; group B, 50 %; group C, 75 %; and group D, 100 % [reproduced from Lidstone et al. (2010) with permission]

maximal reward activity in the brain, e.g., with maximal dopamine-release in subthalamic neurons (Fiorillo et al. 2003).

In the experimental study by Lidstone et al. (2010) only the information about the likelihood of receiving the active drug was varied while in fact all patients received placebo. This resulted in a bell-shaped curve of the placebo response with maximal efficacy in the 50–75 % range, and supports the underlying reward hypothesis (Fig. 6). Scott et al. (2007) found a strong correlation between the placebo effect and rewarding monetary responses: the larger the nucleus accumbens' responses to monetary reward, the stronger the nucleus accumbens' responses to placebos suggesting that placebo responsiveness depends on the functioning and efficiency of the reward system. In this study Scott et al. (2007) used an experimental approach that is typical of clinical trials, i.e., a 50 % chance to receive either placebo or active treatment.

This model can also be used to simulate the results of clinical trials where altered chances to receive active treatment changed the placebo response (see below, Sect. 2.2.4). In this case effective blinding of the investigator may be achieved and may secure unbiased validity of the results. However, it would require substantially more subjects and patients to be studied under both drug and placebo condition and thus may corroborate the intention to mainly study the placebo effect.

### 2.2.4 Inverse Enrichment

Enrichment designs in RCT (as discussed below, Sect. 3.4.1) are chosen to increase the number of patients in the drug arm of the study for ethical reasons (the Declaration of Helsinki requires the least number of patients possible to be included into the placebo arm of studies), for psychological reasons (to improve patient motivation during recruitment), or for methodological reasons (e.g., to test different

drug dosages against one placebo arm). The same strategy can be applied to experimental laboratory studies to enrich the number of volunteers treated with placebo but maintaining the double blinding of the study and avoiding investigator biases.

If for instance, 90 % of volunteers are assigned to placebo and 10 % to a drug, all subjects can still receive and sign the information that they participate in a double-blinded study as long as the true ratio of drug : placebo is not disclosed. This would significantly improve the number of cases available for exploring the placebo response in comparison to a 50:50 balanced chance, and the deception of volunteers is minimized.

## **2.3 Habituation, Sensitization, Learning**

With any repeated measure of any function or symptom in the laboratory or in a RCT, several factors may influence the outcome that are not related to the measure itself but rather to its repetition: extreme values tend to regress towards a mean value over time, participants may learn to distinguish “signals” from “noise” and thereby alter the signal-to-noise ratio of the response, volunteers may habituate to the stimulus, and systems stimulated may either sensitize or desensitize with repetitions. Patients and volunteers may also “learn” what is expected as a response and may want to please the doctor or experimenter (“placebo” in its original meaning as “it may please”). Finally, if intervals between measures are longer, interfering environmental conditions (time, circadian rhythms, other cycles or events) may directly or indirectly influence the measure differentially. In RCT, such influences are taken care of by unbiased randomization of participants into the different study arms, since this warrants an overall averaged effect of all factors in all groups. This holds true also for any spontaneous variation in clinical symptoms over time, as it is the case in many chronic medical conditions (see below, Sect. 3.2.1).

### **2.3.1 “No-Treatment” Controls in the Laboratory**

The equivalent of a “no-treatment” control condition in laboratory experiments is the inclusion of a group in which the experimental measures are taken at the same frequency than in the experimental (placebo) group but without a placebo intervention. Such a “no-treatment” control is usually unblinded (also in RCTs), and subjects are regularly told that they belong to the control group. In RCTs this has substantial effects of the motivation of the patients to continue participation. Whether healthy volunteers in the laboratory respond differently may depend (among other) on the monetary compensation of volunteers, but other effects have never been explored.

Another open question of a “no-treatment” control group in experimental settings is whether and to what degree “no-treatment” implies that not only all timing aspects of the test, but also all experimental procedures except the presumed drug application need to be similar between the placebo and the control group. For

example, in case of a (placebo = NaCl) injection of a presumed analgesic for visceral pain via a constantly running NaCl infusion line (Schmid et al. 2013), it remains to be determined whether the control condition should include the installation of the infusion line or even another NaCl injection that is labeled as placebo. As the purpose of most experiments performed is to elicit maximal placebo response in the experimental group and minimal response in the control group, this may be another source of biases that affect placebo response data as long as they are performed unblinded for the experimenter.

Similarly, the application of an inert skin cream proposed to be a powerful analgesic against experimental pain requires to apply a non-analgesic skin cream in the respective control condition to make measurements comparable otherwise, the skin may respond differentially between two measurements. However, whether volunteers truly believe that they are “controls” rather than experimental subjects has rarely been tested.

Finally, assessing the spontaneous variation of response to an experimental stimulus in “untreated” volunteers is important for the assessment of placebo responsiveness and a placebo responder analysis (as discussed below, Sect. 3.2).

### 2.3.2 Providing Models (Social Learning)

Another systematic way to elicit placebo responses and to control for their efficacy is to use instructed “models” that demonstrate the effectiveness of the procedure applied before the experimental subjects are tested themselves. The clinical equivalence are other patients that report effective treatment by the drug (or the doctor, or the procedure) to other patients prior to their recruitment into a study. It has been noted that “placebo by proxy” (Grelotti and Kaptchuk 2011; Whalley and Hyland 2013) is an almost completely unknown and unexplored effect in RCT, as we will discuss later (Sect. 3.5.2); in experimental settings however, a few studies have demonstrated its efficacy.

Colloca and Benedetti (2009) were the first to show that strong placebo analgesia can be elicited to the same degree than a conditioning procedure when a volunteer was allowed to observe the pain application and reduction by a presumed drug in another person, prior to being tested him- or herself. In a more recent study (Hunter et al. 2013) they also showed that this does not necessarily require the model to be present in the same room, but that a video demonstration may be sufficient, and that empathy with the patient model is not a prerequisite for its efficacy. Others (Swider and Babel 2013; Vögtle et al. 2013) have shown that also strong nocebo effects (hyperalgesia) can be elicited this way, and that (among others) the gender of the model and the experimental subjects determine the efficacy of such modeling.

This raises another relevant issue in experimental setting, especially with respect to pain and placebo analgesia: whether the gender of the experimenter and experimental volunteers play an important role in the response, and to what degree both interact. A number of studies (Aslaksen et al. 2007; Aslaksen and Flaten 2008) have pointed toward such an effect, but data are inconclusive and in part contradictory (Weimer et al. 2010).

Finally, experimental models may also operate without notice of the experimenter: recruitment of experimental subjects often runs by hear-say and subjects informing each other about the options to participate in experiments for monetary reimbursement reasons. It has never been properly assessed whether this takes influence on the experimental findings.

### **2.3.3 Providing Reinforcement (Instrumental Learning)**

Beyond the question whether the mechanisms by which placebo responses occur include social and instrumental learning (and not only Pavlovian conditioning) (which is not the topic of our review) is the fact that providing (monetary) reinforcement for pain-suppressing behavior has been shown to elicit placebo analgesia: when healthy participants were trained to suppress painful mimic expressions during electrical stimulation, they reported lower pain levels compared to baseline stimulations with the same intensity (Kunz et al. 2011).

This calls into question whether many of the procedures installed in placebo research that operate with monetary reward for enduring painful stimuli (at an individually assessed threshold on a visual analog scale) may in fact be biased by indirect reinforcement mechanisms. This could also account for the fact that rather than pain and other sensory thresholds, cognitive assessments of standardized stimuli are responsive to placebo interventions.

## **2.4 Predicting Placebo Responders**

The question whether “placebo responders” (patients and volunteers who reliably respond to a placebo application in a single setting) truly exists has been raised (Kaptchuk et al. 2008) but not answered. Posthoc analyses have been used both for RCT as well as for experimental studies to identify individuals who would show significant responses following a placebo application, with the prediction based on data collected prior to the intervention. The latter requirement is not always met in prediction studies: Definition of a responder based on median split (or any other separation) of the response data (Elsenbruch et al. 2012) is unacceptable, as this is a posthoc selection of the (best) predictor variables selected from a battery of tests installed in the study, thereby creating a strong publication bias. Prediction analysis instead should be based on a multifactorial regression analysis of the entire response range (rather than a dichotomous grouping) within the experimental (placebo) group compared to a “no-treatment” control group.

In a review of the respective literature we (Horing et al. 2014) identified 3 classes of predictor variables: cognitive and motivational predictors (situational optimism, self efficacy, coping strategies), other psychological predictors (suggestibility, bodily self-awareness), and symptom-related predictors (especially with respect to pain and pain control). For a retrospective analysis of own data (Horing 2013) we found the placebo response to be depending on an internal “locus of control,” contrary to common belief: A higher internal locus of control was associated with

lower placebo responsiveness in the experimental group, but with higher responses in the “no-treatment” control group.

However, more questions need to be answered: Are placebo responders responding to the same placebo intervention twice or more? Do placebo responders respond to different placebo interventions across modalities, e.g., in pain studies as well as in studies investigating cognitive responses? Is placebo responsiveness a stable condition over time, and how long can an experimental or clinical placebo response be observed?

Only very few studies have ever shown that placebo response in one study predicts response in a subsequent study, be it within the same domain (Whalley et al. 2008) or across modalities (Kaptchuk et al. 2008). The reason for this paucity of data is obvious: it would require investigation protocols that would exceed (by time, money, organizational efforts, and other determinants) the possibilities of most experimental laboratories.

## 2.5 Avoiding Ethical Conflicts

As discussed above, it cannot be the purpose of a review paper on trial designs to also review and discuss the various ethical aspects that are associated with the use of placebos in experiments, in RCT and in the clinics. However, the use of placebos in experimental research (and *not* in RCTs) raises some specific concerns that need to be addressed here as they have immediate consequences for the conductance of such experiments.

Most experiments that are performed by the majority of placebo researchers imply some type of deception of the volunteers (and in some cases also of the patients) that have stirred discussion about its acceptability (Miller et al. 2005). Different from informed consent in RCT where patients know that they may or may not receive a placebo pill or intervention, in experimental research they are incompletely informed about the purpose of the study and are told instead a “cover story” to hide that the investigation is done to induce a placebo response. Similar to research in lie detection, placebo research may not be able to generate reliable results without the use of deception.

In placebo research, two ethical principles are conflicting: autonomy which requires a fully informed patient and informed consent and assumes full autonomy of the patient, and beneficence which requires optimizing treatment effects and minimizing negative effects, including nocebo effects from informed consent. Many ethical review boards prioritize autonomy and informed consent over beneficence, although this priority should be continuously reevaluated, and new options such as “patient authorized concealments” are to consider.

For experimental research, ethicists have found a similar way out of this dilemma: the introduction of the “authorized deception” (Miller et al. 2005) whereby volunteers in experiments give written informed permission to not being fully informed about the purpose of the study prior to its conductance, to avoid challenging the entire experiment. It has been shown that in comparison to a fully

deceptive study, authorized deception produces similar placebo analgesia with experimental pain in the laboratory (Martin and Katz 2010).

### 2.6 One Size Fits All? The “Free Choice” Paradigm

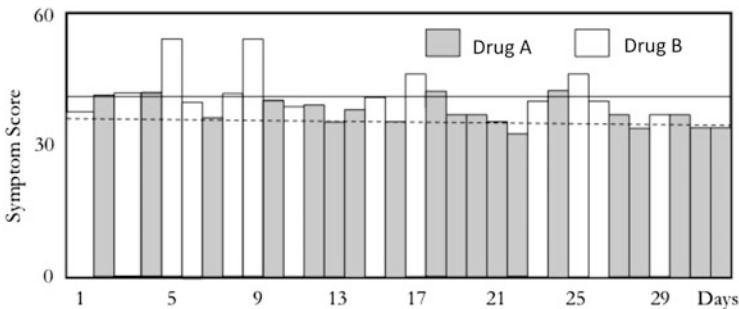
The free-choice paradigm (FCP) most radically breaks with current traditions in clinical and experimental placebo research by introducing the option to choose between drug and placebo to the patient/volunteer (Enck et al. 2012a).

The design allows volunteers/patients to choose between two pills different in colour. They receive the correct information that one contains the drug while the other contains the placebo, but that conditions are double-blinded. In this case no deception is obvious, and hence ethical limitations are minimal, and the dependent variable for measuring drug efficacy is the choice behaviour rather than reported symptoms or symptom improvement.

The design does neither manipulate the information provided to participants and patients, nor does it manipulate the timing of drug release, both of which are common when novel designs are proposed in experimental studies on the placebo effect in healthy volunteers. It thus avoids ethical concerns (deception) in case of inclusion of patients. It also increases the number of events that can be used for evaluation of drug efficacy, e.g., superiority of drug over placebo by computing.

One has, however, to make sure that patients indeed select and do not take both pills simultaneously, thus undermining the intention of the design. It further has to be made sure that technical solutions are installed to warrant appropriate compliance, to prevent over-dosage, and to monitor drug intake.

Other restrictions may be short-acting effects of the drug, the need for steady drug levels, effects on symptoms rather than biochemical disease indicators, hence symptomatic endpoints rather than disease biomarkers. In this case, the primary outcome measure of drug testing is the “selection behavior” of patients (Fig. 7).



**Fig. 7** The “free choice paradigm”: patients can choose daily between drugs A and B. The efficacy measures are either the average symptom score with A (solid line) and B (dotted line) or the number on days with A and B were taken [adopted from Enck et al. (2012a) with permission]

The FCP may be regarded as a modification of the “adaptive response design” (Rosenberger and Lachin 1993), the “early-escape design” (Vray et al. 2004) and other adaptive strategies (Zhang and Rosenberger 2006). It may offer an alternative approach to common drug test procedures, though its statistics have still to be established.

Other requirements of such an approach may be due to the fact that the patient is allowed to switch to the other condition at any time, hence, the pharmacodynamics of the compound under investigation have to be appropriate, e.g., the speed of action, and the feasibility of on-demand medication. It would, on the other hand, allow assessment of drug efficacy via the choice behavior rather than with symptomatic endpoints.

With the FCP, no randomization is needed as all patients have the choice between drug and placebo at predefined time points. Since reasons to alter from 1 day to the next may vary within and across patients, they need to be assessed continuously, e.g., by symptom diaries, and may be taken as covariates in the efficacy analysis. Whether the FCP is suitable for clinical trials in patients needs to be shown in the future.

---

### 3 Clinical Designs to Explore the Placebo Effect

Clinical trials serve a different purpose than most experimental trials: they attempt to demonstrate clinical efficacy of a drug (or any other intervention) *against* a placebo control condition, thus attempting to prove superiority of the therapy under investigation against a placebo condition. In consequence, they try to minimize rather than to maximize (Enck et al. 2013) the placebo response in patients and volunteers. Several design variants have been developed to meet this goal.

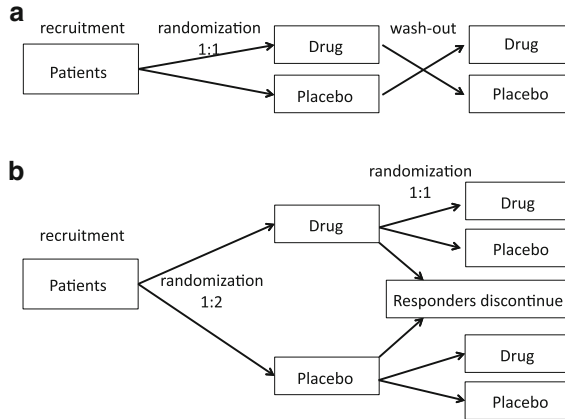
#### 3.1 Identifying Placebo Responders

Ideally, one would wish to identify potential responders to placebo treatment before a study starts, or at least before it is formally evaluated. Any other (posthoc) exclusion of individuals from trial evaluation would be suspected to be severely investigator-biased. Therefore, a number of study designs have been proposed to deal with this issue.

##### 3.1.1 Crossover Designs

From the beginning of RCTs in drug trials in the early 50s and 60s of the last century, it was evident from trial statistics that within-subject variability of responses is lower than between-subject variability under most clinical conditions. In consequence, the idea of each subject providing his/her own control data is at hands and promotes the idea of crossover trials in which patients receive both the drug and placebo in separate phases (with wash-out periods in-between) and in





**Fig. 8** The conventional cross-over design (a) and sequential parallel comparison design (SPCD) according to Fava et al. (2003) (b). Note that randomization schemes may be unbalanced in the RPCD, and that only nonresponders to drug or placebo in Phase 1 are re-randomized to drug or placebo in Phase 2 while responders discontinue. This allows merging of Phase 1 and Phase 2 data in case treatment periods are equally long [see Ivanova et al. (2011) for the statistics]

completely double-blinded randomized and balanced order (Fig. 8a). This was the dominant drug study design in the second half of last century trials.

Crossover trials at the same time support patient recruitment since all patients can be confirmed that at one stage of the study they would receive active treatment. However, by the same mechanisms they encourage patients to compare both treatment phases, and may lead to increased drop-out rates in the second treatment phase if effects and side-effect profiles are so distinct that the switch from drug to placebo discourages continuation. Taken together, crossover designs do not seem to optimize assay sensitivity.

While the risk of un-blinding could be controlled for by using “active placebos” (see below, Sect. 3.3.2) that mimic side-effects of the drug under investigation, crossover trials have also been questioned because treatments in the first phase may generate conditioning effects during the second phase. This has been demonstrated in clinical and experimental studies (e.g., Suchman and Ader 1992; Colloca and Benedetti 2006; Kessner et al. 2013).

### 3.1.2 Placebo and Drug Run-Ins

As a further step in early identification and elimination of placebo responders in drug trials, placebo run-in phases (of days or weeks or even longer) were frequently implemented in RCTs. During this phase all patients receive placebo (and this information was usually provided in the informed-consent information), and those responding with symptom improvement were excluded from the study prior to randomization to drug or placebo.

This pragmatic way of dealing with the placebo response has however two limitations: it assumes that being a placebo responder or a placebo nonresponder

is a stable individual trait that prevents the placebo responses to occur in nonexcluded patients subsequently treated by placebo—which is not the case (Lee et al. 2004). Specifically repeated treatment period designs (see below, Sect. 3.1.4) have demonstrated this effect.

Also, it carries the risk of systematically eliminating an essential subgroup of patients with a specific indication to be excluded from being studied in RCTs, e.g., patients with minor symptom severity that are prone to respond to placebo (Bridge et al. 2009; Kirsch et al. 2008; Enck et al. 2009), although they subsequently may receive the drug prescribed once it is on the market. Such a selection bias needs to be controlled for otherwise drug approval authorities may be inclined to limit the indication for the drug under investigation.

Finally, this design feature is usually nonblinded for the investigator (and maybe for some patients if they read the patient information carefully) and thus generates a bias in clinical assessment.

Drug run-in periods to identify (and exclude) patients that do not respond to the drug at all serve the same purpose of enhancing assay sensitivity, but they run a similar risk: that the drug-responders represent only a subset of all patients with this disease which may invalidate the clinical usefulness of the drug, or its general indication. In addition, especially responders during run-in will notice when they are subsequently randomized to placebo (similar to the effect in crossover trials) and will be unblinded, as will be the treating physician. Drug run-ins will therefore increase the drug effect and decrease the placebo effect, which may be helpful in early phases of drug development only, e.g., for dose-finding.

### 3.1.3 Randomized Run-in/Withdrawal

An elegant and unbiased way to test whether the switch from placebo to drug (run-in) and from drug to placebo (withdrawal) creates strong placebo/nocebo effects is to implement a randomized run-in and withdrawal design (Fig. 9). It is currently favored by US Food and Drug Administration (FDA) and the European Medicinal Agency (EMA), especially with patient reported outcome (PRO) measures.

Days	D1	D2	D3	D4	D5	D6	D7	...	Di	Dj	Dv	Dw	Dx	Dy	Dz
Subj 1	P	D	D	D	D	D	D	...	D	D	P	P	P	P	P
Subj 2	P	P	D	D	D	D	D	...	D	D	D	P	P	P	P
Subj 3	P	P	P	D	D	D	D	...	D	D	D	D	P	P	P
Subj 4	P	P	P	P	D	D	D	...	D	D	D	D	D	P	P
Subj 5	P	P	P	P	P	D	D	...	D	D	D	D	D	D	P
...															
Subj x	P	P	P	P	P	P	P	...	P	P	P	P	P	P	P
Subj y	P	P	P	P	P	P	P	...	P	P	P	P	P	P	P

**Fig. 9** Schematic drawing of the randomized run-in and withdrawal: patients 1–5 start treatment at the same time but receive placebo (P) initially for a variable period of time before being switched to the drug (D) in a double-blinded manner. Similarly, at the end of a set period of the study patients are switched from the drug to placebo at variable time points. Individuals x and y receive placebo throughout the entire study [adopted from Enck et al. (2013) with permission]

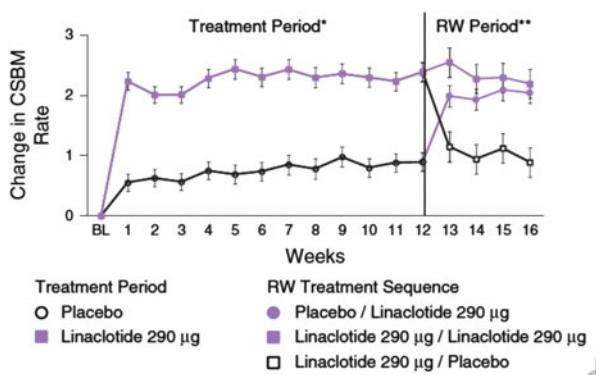
Here the switches from drug to placebo and the drug withdrawal is completely blinded for patients and investigators, and as both are not standardized with respect to timing but may occur within a pre-set time window, symptom improvements (at run-in) and symptom worsening (at withdrawal) may allow the separation of “true” drug responses from drug + placebo compound effects. As this design is rather new, not many data are available to test this hypothesis (Rao et al. 2012).

### 3.1.4 Repetitive Drug Application Phases

A novel strategy that has recently been favored by drug approval authorities in chronic diseases in which cyclic waxing and waning of symptoms is common (such as in irritable bowel syndrome, IBS) is to implement repetitive phases of drug treatment with or without complete re-randomization of patients to drug or placebo, thus going beyond the classical crossover design (see above, Sect. 3.1.1) (Fig. 10). However, this is not primarily to distinguish between drug and placebo response *within* a patient but to demonstrate whether a drug that is taken for some time (and maybe even “on demand,” given the low medication compliance in many chronic conditions) loses or maintains its efficacy during a subsequent treatment period (Rao et al. 2012).

As is evident from the example in Fig. 10, a drug may not lose its potency to improve symptoms in Phase 2, but apparently the pretreatment in Phase 1 with either drug or placebo contributes substantially but differentially to the drug efficacy in Phase 2.

An open question in such a design is whether ethical concerns prohibit a complete re-randomization for Phase 2 and allows that patients that received placebo during Phase 1 may receive placebo also during the second treatment



**Fig. 10** Weekly results for complete spontaneous bowel movement (CSBM) frequency for linaclootide patients compared with placebo patients for each of the 12 treatment-period weeks. During the randomized withdrawal (RW) period patients that had received placebo in the treatment period were switched to linaclootide. As is evident, their symptom improvement is lower than the initial improvement seen during the treatment period, even when the initial drug-placebo difference is counted [Reproduced from Rao et al. (2012) with permission]

phase. The same applies to the following two designs that were specifically developed to overcome the high placebo response rates in recent depression RCTs.

The Sequential Parallel Comparison Design (SPCD) (Fava et al. 2003) consists of two phases: In Phase 1, patients are randomized to receive either drug or placebo in a conventional manner (RCT), but eventually with more patients randomized to placebo (Ivanova et al. 2011). For the second phase, patients in the placebo arm are screened for their response, and nonresponders to placebo will re-randomized to receive either drug or placebo during the second phase of the trial (Fig. 8b).

From the trials currently conducted according to this design (Baer and Ivanova 2013) it is evident that the placebo response is regularly lower in Phase 2 as compared to Phase 1. Statistics (Ivanova et al. 2011) allow either evaluating both phases separately or—given equal treatment duration in both phases—to merge data for a common evaluation.

The Two-way Enrichment Design (TED) (Ivanova and Tamura 2011) is similar but goes one step further: it re-randomizes not only placebo nonresponders but also drug-responders to drug or placebo in Phase 2, this way proposing to enhance the drug response and decrease the placebo response of the complete trial.

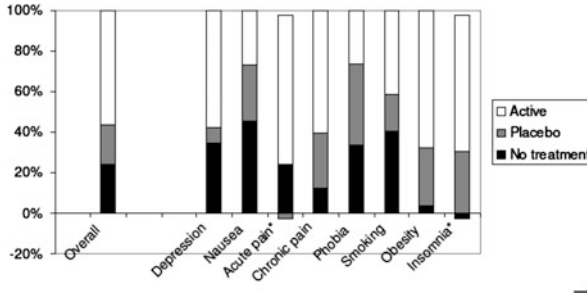
## 3.2 Controlling the Natural Course of Disease

Spontaneous variation of symptoms can occur with all medical conditions, and especially with chronic diseases. They are part of the “unspecific effects” seen in both arms of drug trials (Fig. 1, above). As long as the assumption of “additivity” is correct (Kirsch 2000) such variation may occur in both study arms to the same degree and may therefore be ignored for the evaluation of drug efficacy. However, with the focus on the size and mechanisms of the placebo response in RCTs, this assessment becomes essential to not overestimate the placebo response in clinical trials.

Therefore, “no-treatment” control groups have been mandated by critiques of the current placebo discussion (Hróbjartsson and Gøtzsche 2001, 2004) to account for spontaneous variation of symptoms in many clinical trials that may falsely be attributed to the placebo response. When they meta-analysed studies (Krogsbøll et al. 2009), they found that about half of the placebo response can be attributed to spontaneous remission; this was also true for included pain trials (Fig. 11). They also noted, that the number of studies that used no-treatment controls is low, they are often with benign clinical conditions (smoking cessation, insomnia), and include most often nonmedicinal interventions such as psychotherapy and acupuncture.

### 3.2.1 Waiting Lists, Treatment as Usual

Potential ways around the ethical issue of assigning patients to a “no-treatment” group are waiting list (WL) and “treatment as usual” (TAU) groups that are common control strategies in all nonmedication trials where an inert “placebo” treatment is difficult to provide, such as in psychotherapy, physical rehabilitation,



**Fig. 11** Relative contributions of the spontaneous improvement, effect of placebo, and effect of active treatment to the change from baseline seen in the actively treated group in RCTs with a no-treatment control arm in different clinical conditions [reproduced from Krogsbøll et al. (2009) with permission]

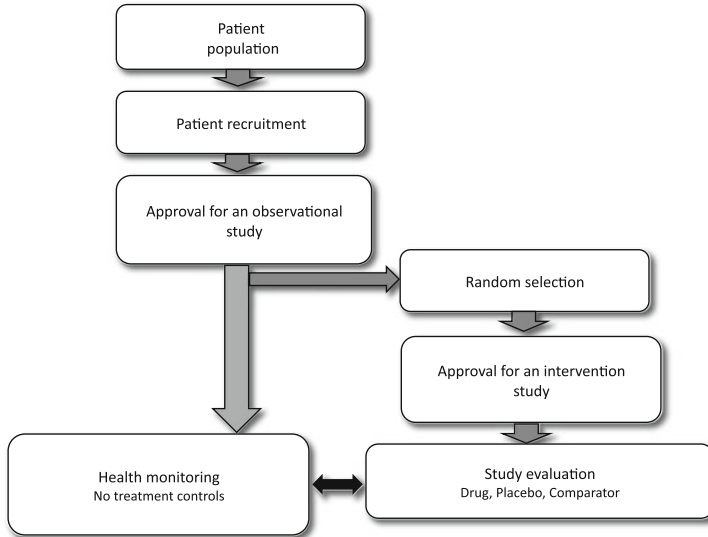
surgery, and “mechanical” interventions (TENS, magnetic stimulation, laser, acupuncture). While some of these therapies have developed their own control strategy (e.g., sham surgery, sham acupuncture), others have relied on WL and TAU. Their limitations are that patients’ expectation to receive effective therapy are at conflict with being randomized to routine treatment (which most of them will have experienced in the past already) and to delays in therapy onset (which may increase the placebo response, but also drop-out rates). This may significantly affect recruitment and compliance in trials, and may lead to biased patients populations in respective studies. A more advanced variant of the WL control strategy is discussed below (Sect. 3.4.3).

WL controls as well as TAU lack credibility as proper control groups in many clinical conditions, and certainly when patients with acute or chronic pain ask for therapy. According to recent meta-analyses (Saarto and Wiffen 2007; Quilici et al. 2009) many drug studies in acute and chronic pain are conducted with comparator drugs rather than with placebos for ethical reasons.

### 3.2.2 The “Zelen Design” or the “Cohort Multiple Randomized Controlled Trial”

A much more acceptable strategy for patients than being randomized into a “no-treatment” control group is the—classical or modified—Zelen design (Zelen 1979) (Fig. 12) that was recently “re-invented” as “cohort multiple randomized controlled trial” (CMRCT) (Relton et al. 2010). It separates recruitment for an observational study that allows assessing spontaneous symptom variation (the “no-treatment” control condition) from randomization for an interventional study, either placebo-controlled or as comparative effectiveness research (CER) study (see below, Sect. 3.6.1).

In this case, the larger the observational cohort the easier the recruitment of a subsample for a treatment study will be: patients are randomly selected from the larger cohort and can be controlled for representativeness, self-selection bias (those that agree to participate in the RCT), and other cohort descriptors.



**Fig. 12** Schematics of the so-called Zelen design (Zelen 1979) that separates recruitment for an observational study from recruitment for one or more intervention studies [adopted from Enck et al (2013) with permission]

However, two limitations apply: the observational cohort needs to be monitored over time (a cross-sectional sample analysis would not be sufficient to account for changes occurring over time), and it needs to be representative for the complete patient cohort affected by the diseases, both in terms of disease features (e.g., symptom severity) as well as disease management (diagnosis, TAU). Once such a cohort is established it may be used for more than one RCT.

### 3.2.3 Registry Trials

Instead of building up an observational cohort for one or more CMRCT, it has recently been proposed to use an already established patient registry that follows a patient cohort (Lauer and D’Agostino 2013). This may be the most elegant way to recruit patients for a trial without randomization into a “no-treatment” control group, but disease registries are only available for a few clinical conditions, e.g., in communicable, in rare, and in the more severe diseases.

## 3.3 Improving Assay Sensitivity

Ways to improve assay sensitivity (the distinction between drug and placebo response in RCTs) include traditional (blinding, active placebos) as well as novel strategies (adaptive designs). We will not discuss here the presumably most important factor in this respect, namely the selection of the primary outcome variable and whether this is a PRO or a disease biomarker.

### 3.3.1 Effective Blinding

While many studies state that they are double-blinded, they rarely report how effective the blinding actually was. In 1986, Ney et al. (1986) stated that the effectiveness of blinding was assessed in less than 5 % of studies conducted between 1972 and 1983. Twenty years later, Hróbjartsson et al. (2007) identified 1,599 blinded randomized studies and found that only 31 (2 %) reported tests for the success of blinding. Even then, only 14 of the 31 studies (45 %) reported that blinding was successful. Ineffective blinding was also noted in pain trials (Machado et al. 2008). Boutron et al. (2006) reviewed methods used in blinding of pharmacological studies and found insufficient report of the efficacy of blinding across studies and conditions. Boehmer and Yong (2009) consequently asked for inclusion of the evaluation of the effectiveness of blinding in RCTs, but this request should also be extended to experimental studies. Blinding in nondrug trials, e.g., in surgery, physical therapy, and with the use of medical devices is even more complicated and potentially costly (Boutron et al. 2007).

A metaanalysis of RCTs in IBS (Shah et al 2013) has recently shown that the drug benefit across 30 trials with 6 groups of drugs is positively and significantly correlated to the number of adverse events reported in the respective drug arm of the trial, indicating a potential un-blinding effect of the adverse events occurring during a trial that co-determines overall drug efficacy. The authors propose that at least presumed treatment allocation should be evaluated after the study.

### 3.3.2 Active Placebos

Active placebos mimic the side effects of a drug under investigation without inducing its main effect in clinical trials. Active placebos in experimental research induce side effects that make the volunteer believe to have received active treatment (e.g., a pain medication); this may be achieved by any perceivable effect following a placebo application, e.g., by skin, olfactory, gustatory, and other signals that are easy to induce and do not interfere with the function under test. Interestingly, active placebos have rarely been used, neither in clinical trials nor in experimental placebo research: Boutron et al. (2006) identified only 6 drug trials with active placebos. Among the few experimental studies that tested active placebos in comparison with inactive ones, Rief and Glombiewski (2012) recently showed that adding a small amount of capsaicin to an otherwise inert nasal placebo spray increased the response rate (placebo analgesia) under a 50:50 chance to that with a 100 % security.

In clinical trials, active placebos are difficult to develop and therefore used only occasionally in a few clinical conditions, e.g., in the treatment of depression (Edward et al. 2005). A Cochrane meta-analysis (Moncrieff et al. 2004) reported only 9 studies with 751 patients with depression, all conducted/ published between 1961 and 1984. In all these cases, the “active placebo” was atropine compared with amitriptylin or imipramine, and all but one study used a parallel-group design. While the overall effect size was in favour of active treatment, it was small compared with placebo-controlled trials using inactive placebos, indicating that

unblinding effects may inflate the efficacy of antidepressants in trials using inert placebos.

### 3.4 Improving Trial Acceptability

Many design features were developed to improve patient recruitment and motivation to participate in drug studies even though they have chances to receive placebo. Patient expectations when enrolled are usually to receive active treatment, and this may lead to discontinuation when the lack of improvement may indicate randomization to placebo (Stone et al. 2005; Lindström et al. 2010).

#### 3.4.1 Unbalanced Randomization

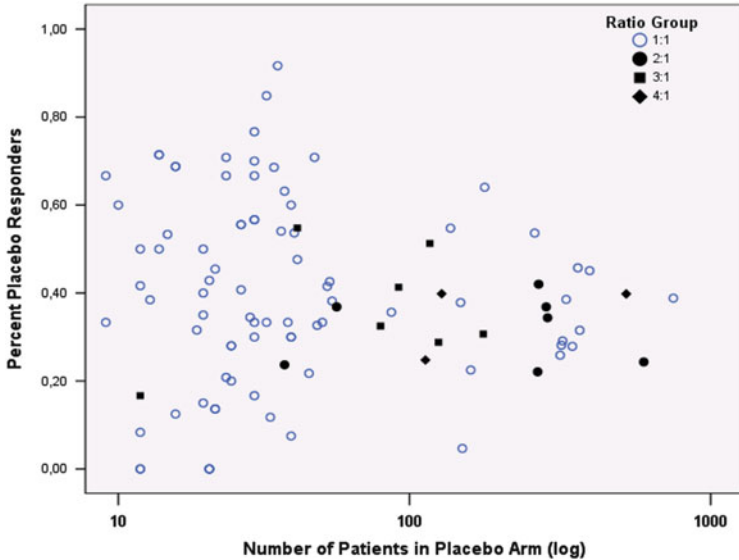
Unbalanced randomization can be used for different purposes: to allow more patients to receive active treatment for ethical reasons, to ease recruitment of patients for practical reasons, or to test more drug doses against a single placebo arm. In all cases, the chances of receiving drug instead of placebo improve.

Experimental evidence shows that the chance of receiving active treatment determines the response to placebo (Lidstone et al. 2010) (see above, Sect. 2.2.3). Clinical data also suggests that the number of study arms in a trial, e.g. with various dosages of the drug against placebo codetermines the size of the placebo and the drug response. In two meta-analyses of depression trials (Papakostas and Fava 2009; Sinyor et al. 2010) it was shown that the lower the likelihood of receiving active treatment (compared to placebo), the lower the response to placebo and to drug. Similar findings were made for migraine (Diener et al. 1999) earlier and for schizophrenia treatment recently (Mallinckrodt et al. 2010): with trial designs that randomized 50 % of patients to either drug or placebo (called 1:1 ratio trials here) the placebo response would be minimal compared to trials with two or more drug arms and higher numbers of patients assigned to active treatment compared to placebo (called 2:1 or  $\geq 2:1$  ratio trials).

Interestingly, this is not supported by data from other areas: Among more than 100 trials with various drugs in irritable bowel syndrome (IBS), 17 used a ratio of drug: placebo greater than 1:1, and these studies yielded a similar placebo response rate than 1:1 studies (Enck et al. 2012b) (Fig. 13).

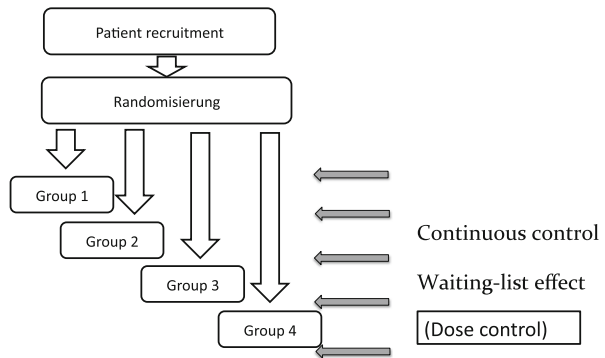
The fact that maximal differences between drug and placebo is achieved with a 1:1 ratio generates an interesting ethical dilemma (Enck et al. 2011b): If exposing patients to placebo carries an ethical burden that requires the minimal number of patients to be assigned to placebo treatment (World Medical Association 2013), more active treatment arms would be in favour. On the other hand, 1:1 trials would require fewer patients to be tested to prove efficacy of the drug over placebo, and thus would claim the same ethical argument to be in favour of 1:1 trials. This dilemma becomes even more virulent with comparator trials (see below, Sect. 3.6.1).





**Fig. 13** Correlation between placebo response rates (%) and number of patients (log transformed) in the placebo arm of 102 randomized, double-blinded placebo-controlled irritable bowel syndrome studies. It is evident that with sample sizes of more than 100 the placebo response tends toward 40 %. Open circles indicate studies powered 1:1 and dark circles indicate studies power  $\geq 2:1$  drug:placebo [adopted from Enck et al. (2012b) with permission]

**Fig. 14** The step-wedge design according to De Allegri et al. (2008) is a modified waiting-list control strategy. Patients are randomized to more than one waiting arm which increases motivation and reduces disappointment, and at the same time allows assessment of a “dose–response” function of waiting for treatment



### 3.4.2 Step-Wedge Design

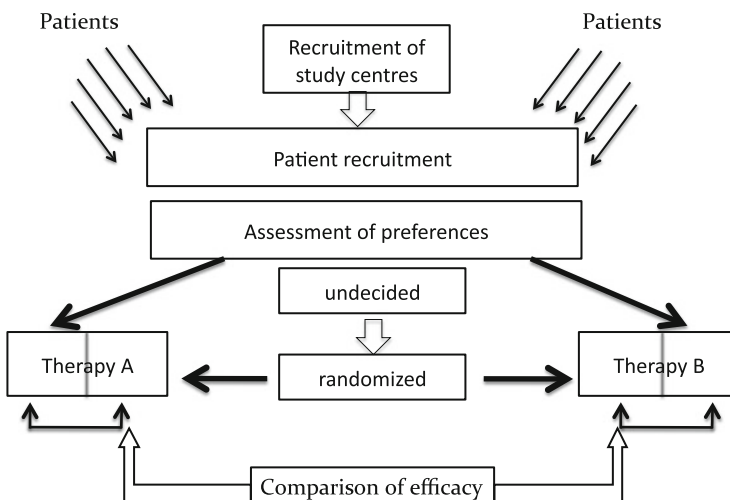
The step-wedge design (De Allegri et al. 2008) is a modification of the WL control group and randomizes patients to different treatment groups that are stacked (immediate begin, begin after  $x$  weeks, after  $y$  weeks, etc.) so that waiting becomes less of a disappointment and waiting time allows assessment of spontaneous variation of symptoms (Fig. 14).

Evidently, the design does not prevent patients from being disappointed to not receive immediate treatment but it minimizes the risk (the more study arms the higher the likelihood to receive earlier treatment) and it allows assessment of a “dose–response” function of waiting.

This latter is of specific interest for a number of reasons: it is known that especially placebo responders in many clinical conditions show lower symptom severity at baseline (Kirsch et al. 2008; Bridge et al. 2009) tend to improve symptoms already during run-in and waiting phases in some conditions (Enck et al. 2009), but not in others (Evans et al. 2004). So far no data exists on the dynamics of waiting effects. In many clinical conditions where no “placebo treatment” is easily available (e.g. in psychotherapy) WL controls are the only option that can be used to control the specificity of therapy. Finally, as discussed above (Sect. 3.2.1), it allows some type of control for spontaneous variation of symptoms under a “no-treatment” control condition, although the expectancy of future treatment may counteract this purpose.

### 3.4.3 Preference Design

Especially under circumstances where more than just one treatment option is available (e.g. psychotherapy versus drug therapy for psychiatric disorders) or in comparator trials (see below, Sect. 3.6.1) where a novel drug is tested against another drug already approved for the same indication instead of being tested against placebo, the “preference design” (King et al. 2005) asks for patients’ preference before patient that do not have any preference are randomized into the treatment arms (Fig. 15).



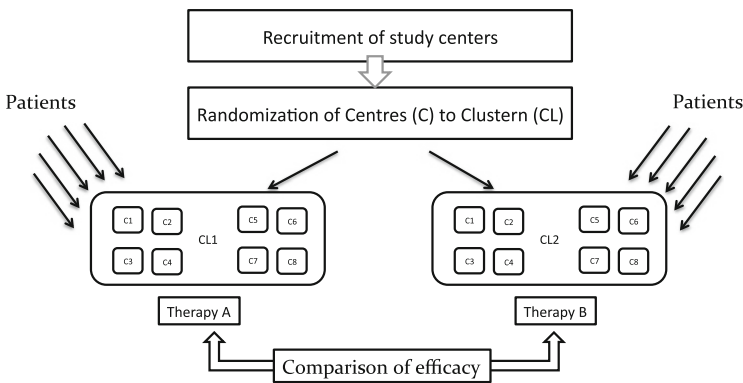
**Fig. 15** The preference design (King et al. 2005) allows patients to chose between alternative treatments when available (e.g., drug vs. psychotherapy) before randomization. It also allows comparison of the efficacy in patient that preferred one arm to patients that were randomized to this arm

Assuming a nearly equal number of patients with preference for one of the two options available, and a substantial number of patients without any preference that will undergo randomization, the preference design would allow assessing whether treatment preference plays a role for treatment outcome by comparing (for each option) the patients that selected the treatment to those that were randomized to the same treatment. This information is usually not available following RCTs but hidden in the efficacy data. The role of preferences can also be included into the overall statistics of comparing both treatment effects. It needs to be shown whether preferences play a role in the placebo response, as has been speculated (Prady et al. 2013).

### 3.4.4 Cluster Randomization

Cluster randomization (Weijer et al. 2012) removes the randomization process further away from the patient: in this case, treatment providers (health care providers, hospitals, private practices) are grouped (clustered) and the decision which cluster provides one therapy and which the other (drug/placebo, drug A/drug B) is randomized (Fig. 16).

In consequence, the patient may not be aware that different treatment options are available, but changing to another cluster is often not feasible due to health care insurance limitations. It has been discussed (McRae et al. 2011) whether such “remote” randomization should be subject to informed consent and that patients should receive the complete information—since they are part of a RCT, ethics approval and patient consent should be identical to conventional trials.



**Fig. 16** Cluster randomization according to Weijer et al. (2012) randomizes treatments to different clusters (CL1, CL2) (treatment centers, hospitals, physicians), while patients are recruited by individual centers (C1 to C8). Thereby, patients have a reduced choice and may not even know that randomization has taken place. This generates ethical issue (McRae et al. 2011)

## 3.5 Developing Individualized Medicine

In our programmatic paper on the future of placebo effects in medicine (Enck et al. 2013) we have argued that for maximizing placebo effects in every-day medicine, individualization of responses to any treatment, including responses to treatment in a RCT—should become the standard in medicine. This includes previous drug history, previous participation in drug trials, and assessment of the role of the social environment of a patient.

### 3.5.1 Previous Drug History

Both positive and negative previous medical experiences co-determine whether a patient is willing to participate in a RCT, and whether or not he/she responds to drug and placebo treatment. It has been shown experimentally that a previous negative (nocebo) experience can affect the degree of placebo analgesia (and hyperalgesia) in experimental pain (Colloca and Benedetti 2006), and that with repetitive exposure to the same placebo analgesia experience can provide long-lasting efficacy (Kessner et al. 2013).

In clinical trials the situation is similar: In Parkinsons' Disease previous experience with a drug for restless leg syndrome determined similar efficacy of the same drug in a subsequent trial; unfortunately, in a second trial this was not the case but rather the opposite happened (de la Fuente-Fernández 2012).

Similar data are available for only a few other clinical conditions (Iovieno and Papakostas 2012), and the current state of knowledge is rather poor. One legal restriction that applies here is that individualized patient data that have been generated in one RCT cannot easily be transferred to another RCT especially when different investigators or drug companies are involved, for protection of the patient's anonymity. A way out of this dilemma could be the organization of a patient registry for RCTs (see below, Sect. 3.5.3).

### 3.5.2 "Placebo by Proxy"

The phenomenon of "placebo by proxy" has been established in assessing the determinants of placebo responses in children: While we know that placebo responses overall are larger in RCT in children and adolescents than in adults (Weimer et al. 2013a), little is known about the underlying mechanisms. Apparent mechanisms that account for high placebo response rates in adult disorders, e.g., the number and intensity of doctor visits during a RCT are not operating in children (Rutherford et al. 2011).

It has been argued (Lewis et al 2005) that placebos could operate by producing changes in how caregivers perceive children symptom changes. Placebos could also operate by producing changes in how caregivers behave toward children, which in turn produce behavioral changes in the child. The concept of "placebo by proxy" has recently received attention both from a methodological point of view (Grelotti and Kaptchuk 2011) as well as in an observational study on temper tantrums in children (Whalley and Hyland 2013).

Grelotti and Kaptchuk (2011) argued that—not only in children—the expectations of a patient towards his/her treatment is based not only on own experience and hopes, but occurs in a social context where proxies (family members, caregivers, relatives) respond to symptoms and their improvement and worsening as well. Because these can exist independently of any placebo response of the patient, their contribution to the patient's response are largely unknown and uninvestigated. One of the paradigmatic examples the authors cite refers to the fact that antibiotics are frequently overprescribed specifically in children because of parents' concerns and wishes (Mangione-Smith et al. 1999). Proxies' influences on (placebo) responsiveness may also be responsible for differences in efficacy reports seen between doctor and patient-reported outcomes, especially in depression (Rief et al. 2009).

Whalley and Hyland (2013) take the argument that placebo by proxy may play an important role especially in children one step further: They investigated whether the efficacy of an impure placebo (Bach flower therapy, a homeopathic remedy) to improve symptoms of temper tantrums in 2–5-year old children would be affected by the parents' beliefs and mood. To exclude any direct effect of physician-child and physician-parent interaction, an automated telephone system was used for symptom recording. The authors found a sustained and significant improvement of tantrum frequency and severity that was strongly correlated to parents' mood. As this was an observational study, the authors cannot conclude on the true nature of the symptomatic improvement but assume that these are “pure” placebo effects. Whether symptom improvements were mirrored in children's behavioral changes or only in parents' perception cannot be concluded from the data.

However, as discussed above not only children but most adult patients have a social environment (family, relatives, friends) that participated in the illness history, is involved in its current care and is interested in its future development. Not only the patients own experience with drugs, but also the experience of these “significant others” may co-determine responses to drug and placebo in a RCT. This field of “placebo by proxy” in adulthood is and remains vastly unexplored as long as reliable methods of assessment are missing.

### 3.5.3 Patient Registry

We have recently argued (Enck et al. 2013) that individualized medicine with respect to placebo responses would require some type of patient registry that serves a dual purpose: protecting patients' anonymity and data collected during one RCT but at the same time make these data available for evaluation of another RCT in which the patient may participate in the future. The legal and ethical rules of such data transfer still need to be established.

This goes far beyond what is current practice in either disease-specific databases (e.g., “. . . to develop a comprehensive database of individuals who are diagnosed with . . . , to better understand the characteristics of these diseases, to determine areas that need further research, and to help pharmaceutical companies with the development of treatments to improve the lives of those affected” (<https://connect.patientcrossroads.org/?org=apfed>) or in databases for drug companies helping

them to evaluate RCT outcomes (Electronic Medical Records), and it also is more than just a recruitment basis for future RCTs to ensure that only properly diagnosed patients are included into such studies.

### 3.6 Dismissing Placebos in RCTs

While placebo-controlled RCT are still regarded as the gold-standard in the development of novel drug treatments, they have come into question for several reasons: the recently released updated version of the Declaration of Helsinki of World Medical Association (WMA) (World Medical Association 2013) calls for an even more restrictive use of placebo controlled trials in drug development, and some countries have banned the use of such trials entirely (Ehni and Wiesing 2008). In consequence, drug approval authorities such as the FDA and the EMA favor head-to-head comparison (also called “comparator trials” or “comparative effectiveness research,” CER) of novel compounds against drugs already marketed for both ethical reasons (no patient without active treatment) as well as economic reasons (novel drugs should be at least equal to what is already available).

#### 3.6.1 Head-to-Head Trials and CER

It is said that CER trials more closely mimic the situation occurring in medical routine where several drugs are available to treat one condition, and where direct comparison of their efficacy is feasible. In contrast, the clinical equivalence of placebo treatment is said to be a “watchful waiting” decision (Hegerl and Mergl 2010) although (as we have discussed above, Sect. 3.2.1) waiting lists are inappropriate control conditions for what happened without treatment.

Because the placebo response is immanent in all medical treatments, *not* applying placebos in RCTs does not result in no placebo response at all but rather in its ignoring during evaluation of the data. As we know from the evaluation of enrichment trials and unbalanced randomization in experiments (see above, Sect. 2.2.3) and in clinical trials (see above, Sect. 3.4.1), providing a 100 % chance to receive active treatment increases the response to both drug and placebo compared to a 50:50 chance as in placebo-controlled trials. However, CER trials lack the direct possibility to assess the placebo response.

In a meta-analytic comparison of CER trials and placebo-controlled trials of the same drugs for treatment of depression it was shown that CER trials enhance the drug response (compared with placebo controlled trials of the same compounds) solely by the expectation to receive a drug by 100 %, and add another 15 % placebo response to the already established average of 40 % from placebo-controlled drug trials for depression (Rutherford et al 2009). Similar data have been shown for CER in schizophrenia (Woods et al. 2005).

This creates an ethical dilemma already discussed above (Sect. 3.4.1): CER trials need up to four times more patients for a statistical test of “noninferiority” than conventional placebo-controlled trials (Leon 2012) which contradicts the statement that the least number of patients should be included in RCT. CER trials are also

associated with substantial increased costs of trials, specifically if the selected appropriate comparator drug requires it to be produced (because it is the property of a competing company), and the provision of double-dummy technology (Marušić and Ferenčić 2013). Finally, the selection of the comparator may force substantial methodological considerations and concerns, if more than one potential comparator is available on the market (Estellat and Ravaud 2012; Dunn et al. 2013).

### 3.6.2 Historic Controls

A 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 (total:  $n = 24,581$  studies) for studies where patients were recruited into a placebo arm of pain trials ( $n = 3,119$ ). After screening and merging of the data (that were 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. It is proposed to use these historic controls as a database rather than recruiting future patients into placebo arms of RCTs with novel compounds. Feasibility of such an approach needs however still to be verified prospectively.

---

## 4 Summary

As we have discussed, both experimental and clinical study designs have attempted to identify placebo responders, to characterize them, and to limit the effects of placebo application of primary and secondary outcome measures, with variable success. Among the different strategies chosen, early identification and exclusion of placebo responders and drug nonresponders seem most promising but carry the risk of selective indication. Enrichment strategies to enhance the placebo–drug difference are most promising for drug development, but for the purpose of characterizing mechanisms of the placebo response, it is most important to distinguish the placebo response from other influences on trial outcomes, especially of spontaneous symptom variation, statistical errors, and response biases. Novel strategies include the use of randomized run-in and withdrawal periods, historic controls, and e-patients but most of them still have to be evaluated.

**Acknowledgment** Supported by a grant from Deutsche Forschungsgemeinschaft

---

## References

- Amanzio M, Pollo A, Maggi G et al (2001) Response variability to analgesics: a role for non-specific activation of endogenous opioids. *Pain* 90:205–215
- Aslaksen PM, Flaten MA (2008) The roles of physiological and subjective stress in the effectiveness of a placebo on experimentally induced pain. *Psychosom Med* 70:811–818
- Aslaksen PM, Myrbakk IN, Høifødt RS et al (2007) The effect of experimenter gender on autonomic and subjective responses to pain stimuli. *Pain* 129:260–268

- Baer L, Ivanova A (2013) When should the sequential parallel comparison design be used in clinical trials? *Clin Invest* 3:823–833
- Benedetti F, Maggi G, Lopiano L et al (2003) Open versus hidden medical treatments: the patient's knowledge about a therapy affects the therapy outcome. *Prevention Treatment* 6(1). <http://psycnet.apa.org/journals/pre/6/1/1a.html>. Accessed 28 Jan 2014
- Benedetti F, Carlino E, Pollo A (2011) Hidden administration of drugs. *Clin Pharmacol Ther* 90:651–661
- Bingel U, Wanigasekera V, Wiech K et al (2011) The effect of treatment expectation on drug efficacy: imaging the analgesic benefit of the opioid remifentanyl. *Sci Transl Med* 3:70ra14
- Boehmer J, Yong P (2009) How well does blinding work in randomized controlled trials?: a counterpoint. *Clin Pharmacol Ther* 85:463–465
- Boutron I, Estellat C, Guittet L et al (2006) Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med* 3:e425
- Boutron I, Guittet L, Estellat C et al (2007) Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 4:e61
- Bridge JA, Birmaher B, Iyengar S et al (2009) Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *Am J Psychiatry* 166:42–49
- Colloca L, Benedetti F (2006) How prior experience shapes placebo analgesia. *Pain* 124:126–133
- Colloca L, Benedetti F (2009) Placebo analgesia induced by social observational learning. *Pain* 144:28–34
- Colloca L, Lopiano L, Lanotte M et al (2004) Overt versus covert treatment for pain, anxiety, and Parkinson's disease. *Lancet Neurol* 3:679–684
- D'Agostino RB (2009) The delayed-start study design. *N Engl J Med* 361:1304–1306
- De Allegri M, Pokhrel S, Becher H et al (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-Fernández R (2012) The powerful pre-treatment effect: placebo responses in restless legs syndrome trials. *Eur J Neurol* 19:1305–1310
- Desai JR, Bowen EA, Danielson MM, Allam RR et al (2013) Creation and implementation of a historical controls database from randomized clinical trials. *J Am Med Inform Assoc* 20:e162–e168
- Diener HC, Dowson AJ, Ferrari M et al (1999) Unbalanced randomization influences placebo response: scientific versus ethical issues around the use of placebo in migraine trials. *Cephalalgia* 19:699–700
- Dunn AG, Mandl KD, Coiera E et al (2013) The effects of industry sponsorship on comparator selection in trial registrations for neuropsychiatric conditions in children. *PLoS ONE* 8:e84951
- Edward SJ, Stevens AJ, Brauholtz DA et al (2005) The ethics of placebo-controlled trials: a comparison of inert and active placebo controls. *World J Surg* 29:610–614
- Ehni HJ, Wiesing U (2008) International ethical regulations on placebo-use in clinical trials: a comparative analysis. *Bioethics* 22:64–74
- Elsenbruch S, Kotsis V, Benson S et al (2012) Neural mechanisms mediating the effects of expectation in visceral placebo analgesia: an fMRI study in healthy placebo responders and nonresponders. *Pain* 153:382–390
- Enck P, Vinson B, Malfertheiner P et al (2009) Placebo effects in functional dyspepsia – reanalysis of trial data. *Neurogastroenterol Motil* 21:370–377
- Enck P, Klosterhalfen S, Zipfel S (2011a) Novel study designs to investigate the placebo response. *BMC Med Res Methodol* 11:90
- Enck P, Klosterhalfen S, Weimer K et al (2011b) The placebo response in clinical trials: more questions than answers. *Philos Trans R Soc Lond B Biol Sci* 366:1889–1895
- Enck P, Grundy D, Klosterhalfen S (2012a) A novel placebo-controlled clinical study design without ethical concerns – the free choice paradigm. *Med Hypotheses* 79:880–882
- Enck P, Horing B, Weimer K et al (2012b) Placebo responses and placebo effects in functional bowel disorders. *Eur J Gastroenterol Hepatol* 24:1–8



- Enck P, Bingel U, Schedlowski M et al (2013) The placebo response in medicine: minimize, maximize or personalize? *Nat Rev Drug Discov* 12:191–204
- Estellat C, Ravaud P (2012) Lack of head-to-head trials and fair control arms: randomized controlled trials of biologic treatment for rheumatoid arthritis. *Arch Intern Med* 172:237–244
- Evans KR, Sills T, Wunderlich GR et al (2004) Worsening of depressive symptoms prior to randomization in clinical trials: a possible screen for placebo responders? *J Psychiatr Res* 38:437–444
- Fava M, Evins AE, Dorer DJ et al (2003) The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom* 72:115–127
- Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299:1898–1902
- Gracely RH, Dubner R, Wolskee PJ et al (1983) Placebo and naloxone can alter post-surgical pain by separate mechanisms. *Nature* 306:264–265
- Grelotti DJ, Kaptchuk TJ (2011) Placebo by proxy. *BMJ* 343:d4345
- Hegerl U, Mergl R (2010) The clinical significance of antidepressant treatment effects cannot be derived from placebo-verum response differences. *J Psychopharmacol* 24:445–448
- Horing B (2013) Placebo effects and their prediction across multiple experimentally induced symptoms: motion sickness, cutaneous heat and cold pain, and rectal distension. Ph.D. thesis, University of Tübingen Medical School, Tübingen
- Horing B, Weimer K, Muth ER et al (2014) Prediction of placebo responses: a systematic review of the literature. *Front Psychol* (in press)
- Hróbjartsson A, Gøtzsche PC (2001) Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med* 344:1594–1602
- Hróbjartsson A, Gøtzsche PC (2004) Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. *J Intern Med* 256:91–100
- Hróbjartsson A, Forfang E, Haahr MT et al (2007) Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol* 36:654–663
- Hunter T, Siess F, Colloca L (2013) Socially induced placebo analgesia: a comparison of a pre-recorded versus live face-to-face observation. *Eur J Pain* 18:914–922. doi:[10.1002/j.1532-2149.2013.00436.x](https://doi.org/10.1002/j.1532-2149.2013.00436.x)
- Iovieno N, Papakostas GI (2012) Does the presence of an open-label antidepressant treatment period influence study outcome in clinical trials examining augmentation/ combination strategies in treatment partial responders/nonresponders with major depressive disorder? *J Clin Psychiatry* 73:676–683
- Ivanova A, Tamura RN (2011) A two-way enriched clinical trial design: combining advantages of placebo lead-in and randomized withdrawal. *Stat Methods Med Res.* doi:[10.1177/0962280211431023](https://doi.org/10.1177/0962280211431023)
- Ivanova A, Qaqish B, Schoenfeld DA (2011) Optimality, sample size, and power calculations for the sequential parallel comparison design. *Stat Med* 30:2793–2803
- Kaptchuk TJ, Kelley JM, Deykin A et al (2008) Do “placebo responders” exist? *Contemp Clin Trials* 29:587–595
- Kelemen WL, Kaighobadi F (2007) Expectancy and pharmacology influence the subjective effects of nicotine in a balanced-placebo design. *Exp Clin Psychopharmacol* 15:93–101
- Kessner S, Wiech K, Forkmann K et al (2013) The effect of treatment history on therapeutic outcome: an experimental approach. *JAMA* 173:1468–1469
- King M, Nazareth I, Lampe F et al (2005) Conceptual framework and systematic review of the effects of participants’ and professionals’ preferences in randomised controlled trials. *Health Technol Assess* 9:1–186, iii–iv
- Kirsch I (2000) Are drug and placebo effects in depression additive? *Biol Psychiatry* 47:733–735
- Kirsch I (2005) Placebo psychotherapy: synonym or oxymoron? *J Clin Psychol* 61:791–803

- Kirsch I, Deacon BJ, Huedo-Medina TB et al (2008) Initial severity and antidepressant benefits: a meta-analysis of data submitted to the food and drug administration. *PLoS Med* 5:e45
- Krogsbøll LT, Hróbjartsson A, Gøtzsche 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
- Kunz M, Rainville P, Lautenbacher S (2011) Operant conditioning of facial displays of pain. *Psychosom Med* 73:422–431
- Lauer MS, D'Agostino RB Sr (2013) The randomized registry trial - the next disruptive technology in clinical research? *N Engl J Med* 369:1579–1581
- Lee S, Walker JR, Jakul L, Sexton K (2004) Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. *Depress Anxiety* 19:10–19
- Leon AC (2012) Challenges in designing comparative-effectiveness trials for antidepressants. *Clin Pharmacol Ther* 91:165–167
- Levine JD, Gordon NC, Fields HL (1978) The mechanism of placebo analgesia. *Lancet* 2:654–657
- Lewis DW, Winner P, Wasiewski W (2005) The placebo responder rate in children and adolescents. *Headache* 45:232–239
- Lidstone SC, Schulzer M, Dinelle K et al (2010) Effects of expectation on placebo-induced dopamine release in Parkinson disease. *Arch Gen Psychiatry* 67:857–865
- Lindström D, Sundberg-Petersson I, Adami J et al (2010) Disappointment and drop-out rate after being allocated to control group in a smoking cessation trial. *Contemp Clin Trials* 31:22–26
- Lund K, Vase L, Petersen GL et al (2014) Randomised controlled trials may underestimate drug effects: balanced placebo trial design. *PLoS ONE* 9:e84104
- Machado LA, Kamper SJ, Herbert RD et al (2008) Imperfect placebos are common in low back pain trials: a systematic review of the literature. *Eur Spine J* 17:889–904
- Mallinckrodt CH, Zhang L, Prucka WR et al (2010) Signal detection and placebo response in schizophrenia: parallels with depression. *Psychopharmacol Bull* 43:53–72
- Mangione-Smith R, McGlynn EA, Elliott MN et al (1999) The relationship between perceived parental expectations and pediatrician antimicrobial prescribing behavior. *Pediatrics* 103:711–718
- Martin AL, Katz J (2010) Inclusion of authorized deception in the informed consent process does not affect the magnitude of the placebo effect for experimentally induced pain. *Pain* 149:208–215
- Marušić A, Ferenčić SF (2013) Adoption of the double dummy trial design to reduce observer bias in testing treatments. *J R Soc Med* 106:196–198
- McRae AD, Weijer C, Binik A et al (2011) When is informed consent required in cluster randomized trials in health research? *Trials* 12:202
- Metrik J, Rohsenow DJ, Monti PM et al (2009) Effectiveness of a marijuana expectancy manipulation: piloting the balanced-placebo design for marijuana. *Exp Clin Psychopharmacol* 17:217–225
- Miller FG, Wendler D, Swartzman LC (2005) Deception in research on the placebo effect. *PLoS Med* 2:e262
- Moncrieff J, Wessely S, Hardy R (2004) Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev*:CD003012
- Ney PG, Collins C, Spensor C (1986) Double blind: double talk or are there ways to do better research? *Med Hypotheses* 21:119–126
- 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* 19:34–40
- Petrovic P, Kalso E, Petersson KM et al (2002) Placebo and opioid analgesia: imaging a shared neuronal network. *Science* 295:1737–1740
- Pollo A, Torre E, Lopiano L et al (2002) Expectation modulates the response to subthalamic nucleus stimulation in Parkinsonian patients. *Neuroreport* 13:1383–1386

- Prady SL, Burch J, Crouch S et al (2013) Insufficient evidence to determine the impact of patient preferences on clinical outcomes in acupuncture trials: a systematic review. *J Clin Epidemiol* 66:308–318
- Quilici S, Chancellor J, Löthgren M et al (2009) Meta-analysis of duloxetine vs. pregabalin and gabapentin in the treatment of diabetic peripheral neuropathic pain. *BMC Neurol* 9:6
- Rao S, Lembo AJ, Shiff SJ et al (2012) A 12-week, randomized, controlled trial with a 4-week randomized withdrawal period to evaluate the efficacy and safety of linaclotide in irritable bowel syndrome with constipation. *Am J Gastroenterol* 107:1714–1724
- Relton C, Torgerson D, O’Cathain A et al (2010) Rethinking pragmatic randomised controlled trials: introducing the “cohort multiple randomised controlled trial” design. *BMJ* 340:e1066
- Rief W, Glombiewski JA (2012) The hidden effects of blinded, placebo-controlled randomized trials: an experimental investigation. *Pain* 153:2473–2477
- Rief W, Nestoriuc Y, Weiss S et al (2009) Metaanalysis of the placebo response in antidepressant trials. *J Affect Disord* 18:1–8
- Rief W, Bingel U, Schedlowski M et al (2011) Mechanisms involved in placebo and nocebo responses and implications for drug trials. *Clin Pharmacol Ther* 90:722–726
- Rosenberger WF, Lachin JM (1993) The use of response-adaptive designs in clinical trials. *Contemp Clin Trials* 14:471–484
- Rutherford BR, Sneed JR, Rosse SP (2009) Does study design influence outcome? *Psychother Psychosom* 78:172–181
- Rutherford BR, Sneed JR, Tandler JM et al (2011) Deconstructing pediatric depression trials: an analysis of the effects of expectancy and therapeutic contact. *J Am Acad Child Adolesc Psychiatry* 50:782–795
- Saarto T, Wiffen PJ (2007) Antidepressants for neuropathic pain. *Cochrane Database Syst Rev*: CD005454
- Schmid J, Theysohn N, Gaß F et al (2013) Neural mechanisms mediating positive and negative treatment expectations in visceral pain: a functional magnetic resonance imaging study on placebo and nocebo effects in healthy volunteers. *Pain* 154:2372–2380
- Scott DJ, Stohler CS, Egnatuk CM et al (2007) Individual differences in reward responding explain placebo-induced expectations and effects. *Neuron* 55:325–336
- Shah E, Triantafyllou K, Hana AA et al (2013) Adverse events appear to unblind clinical trials in irritable bowel syndrome. *Neurogastroenterol Motil* 26:482–488. doi:10.1111/nmo.12289
- Sinyor M, Levitt AJ, Cheung AH et al (2010) Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. *J Clin Psychiatry* 71:270–279
- Stone DA, Kerr CE, Jacobson E et al (2005) Patient expectations in placebo-controlled randomized clinical trials. *J Eval Clin Pract* 11:77–84
- Suchman AL, Ader R (1992) Classic conditioning and placebo effects in crossover studies. *Clin Pharmacol Ther* 52:372–377
- Swider K, Babel P (2013) The effect of the sex of a model on nocebo hyperalgesia induced by social observational learning. *Pain* 154:1312–1317
- Vase L, Riley JL, Price DD (2002) A comparison of placebo effects in clinical analgesic trials versus studies of placebo analgesia. *Pain* 99:443–452
- Vögtle E, Barke A, Kröner-Herwig B (2013) Nocebo hyperalgesia induced by social observational learning. *Pain* 154:1427–1433
- Volkow ND, Wang GJ, Ma Y et al (2003) Expectation enhances the regional brain metabolic and the reinforcing effects of stimulants in cocaine abusers. *J Neurosci* 23:11461–11468
- Vray M, Girault D, Hoog-Labouret N et al (2004) Methodology for small clinical trials. *Therapie* 59:273–286
- Weijer C, Grimshaw JM, Eccles MP et al (2012) The Ottawa statement on the ethical design and conduct of cluster randomized trials. *PLoS Med* 9:e1001346
- Weimer K, Enck P, Klosterhalfen S (2010) Gender effects in placebo responses. *Z Med Psychol* 19:146–153

- Weimer K, Gulewitsch M, Schlarb AA et al (2013a) Placebo effects in children: a review. *Pediatr Res* 74:96–102
- Weimer K, Horing B, Stürmer J et al (2013b) Nicotine expectancy differentially affects reaction time in healthy non-smokers and smokers depending on gender. *Exp Clin Psychopharmacol* 21:181–187
- Weimer K, Horing B, Walentin S et al (2013c) Nicotine stimulus expectancy but not nicotine affects cognitive performance in healthy smokers and non-smokers. *Psychosom Med* 75:A115
- Whalley B, Hyland ME (2013) Placebo by proxy: the effect of parents' beliefs on therapy for children's temper tantrums. *J Behav Med* 36:341–346
- Whalley B, Hyland ME, Kirsch I (2008) Consistency of the placebo effect. *J Psychosom Res* 64:537–541
- Woods SW, Gueorguieva RV, Baker CB et al (2005) Control group bias in randomized atypical antipsychotic medication trials for schizophrenia. *Arch Gen Psychiatry* 62:961–970
- World Medical Association (2013) World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310:2191–2194
- Zelen M (1979) A new design for randomized clinical trials. *N Engl J Med* 300:1242–1245
- Zhang L, Rosenberger WF (2006) Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics* 62:562–569