Myths and Facts about Placebo

Hans-Joachim Kremer*

Medical Writing Service, Alemanennstraße 101, 79117 Freiburg, Germany

Abstract: Many patients, physicians, and sometimes even academics have questionable perceptions of placebo and the so-called placebo effect, Many believe that placebo have its own effects. Although psychological aspects, namely expectations of patients or the persuasive power of the physicians, might sometimes be substantial, such aspects may have little or even no relevance in other situation where placebo control is essential nevertheless. Even in settings where psychological effects should be envisaged, their extent is usually highly variable, indicating that other factors might still exceed the importance of psychological effects. Placebo is defined in US regulations as an inactive preparation designed to resemble the test drug as far as possible. This means that placebo itself cannot be effective. If it would, its correctness is challenged and it should be replaced if still possible. And as placebo is not effective, it can also not have secondary effects, *vulgo* side effects. Placebo is always used for two reasons: To control bias and to provide the reasonably largest delta, i.e. the difference between two treatments. Placebo should never be interpreted as being able to cause effects.

Keywords: Placebo-effect, definition, bias, delta, nocebo.

1. INTRODUCTION

Those who are working in the area of clinical research and development for pharmaceutical companies are often confronted by patients and clinicians with surprising and sometimes questionable perceptions with respect to our daily work. Often it is stated something like: Placebo works in some indications and under certain circumstances; placebo acts in this or that way; placebo has plenty of effects; it might be sufficient to treat patients with placebos, just because so many and, at a glance, substantial effects have been reported for the placebo groups from clinical trials. And in general, the physicians' suggestive actions, their care, and their devotion to the patients would be so strong that this alone could cure and help many patients.

It might be acceptable, if patients or laypersons share such perceptions, but misperceptions appear to be common also among physicians and even researchers, vice versa, enforcing those myths. Many researchers might confess that placebo does not exert objective, but subjective effects. Some researchers even claimed to have demonstrated objective effects of placebo. There might be some evidence supporting the suggestive power of physicians to cure and help patients on average. Instead, most of the "effects" accounted to placebo itself are rather attributable to other causes and do usually not require psychological explanations, at least if blinding could be guaranteed during the whole trial.

Therefore, I wrote this article to challenge the myths and to explain the facts about placebo and the so-called placebo-effect.

2. DEFINITION OF PLACEBO

My search for a definition of placebo started with surprises. Most patients and physicians might use Wikipedia articles, at least for a first look; so did I. The English article on placebo provides, after a "has been defined", the phrase "a substance or procedure… that is objectively without specific activity for the condition being treated". Then the article presents a reference [1] for this definition, however, that paper was in fact on regression to the mean and actually warned of interpreting patient improvements as causal effects; whatsoever, that paper does not appear to be a reasonable reference for such a definition. The German Wikipedia article provides another definition (essentially: "without pharmacologically active ingredients"), but without a reference; at least the closest references [2] does not provide this definition. Then, the article continues claiming a differentiation between various kinds of placebo: "True or pure placebos, active placebos, pseudoplacebos"; for these the reference might be useful. The French Wikipedia article provides a similar definition as the German, again without reference. The reader may wonder here, whether it is worthwhile for the present paper criticising Wikipedia articles. This is certainly not my main intention, but these circumstances may illustrate the scientific and public confusion on basic terms.

Although the area of clinical trials is highly regulated with many international guidelines, the search for an internationally accepted definition of placebo turned out

*Address correspondence to this author at the Medical Writing Service, Alemanennstraße 101, 79117 Freiburg, Germany; Tel: +49 761 6966 480; E-mail: hans-joachim.kremer@t-online.de

ISSN: 2223-3806 / E-ISSN: 1927-5951/16

a bit disappointing; the next surprise. In fact, neither ICH[1] E6 (Good Clinical Practice, GCP [3]), nor E8 (General Considerations for Clinical Trials [4]), nor E9 (Statistical Considerations for Clinical Trials [5]) provide a definition for placebo, and this although E6 as well as E9 provide otherwise many definitions for important terms used in clinical research. At least the E10 (Choice of Control Groups [6]) provides some thoughts that are very similar to the US definition outlined below. Among the regulatory rules in the European Union, I could also not identify a definition; neither the old European GCP Directive (2001/20/EC [7]) nor the forthcoming EU Clinical Trials Regulation 536/2014 [8] provide a definition for placebo.

It appears that the only definition of placebo issued by legislators or regulators is contained in the basic US legislation Code of Federal Regulations (CFR), namely in 21 CFR 314, §126, b 2 i:

> *Placebo concurrent control: The test drug is compared with an inactive preparation designed to resemble the test drug as far as possible.*

This definition appears outstanding, not only because of its legal source. It is clear and simple. It stresses the essential aspect, *an inactive preparation*, without any modifying attributes such as "pharmacologically" or claiming that the feature *inactive* might be confined to "objectively" proven inactivity, or that the inactivity might be confined to "specific" aspects. Just inactive! Moreover, it stresses the other essential aspect, namely the maximum similarity to the test drug.

Clearly, all US studies using placebos should refer to the 21 CFR 314 definition as it is law in the USA. In a globalised world, any non-US company would be badly advised if deviating from this rule, making this definition in fact globally valid. Even more disturbing is that many researchers and Wikipedia authors appear to be ignorant of this definition.

While not marked explicitly as definition, the ICH E10 guideline [6] states that placebo is expected to be to "an identical-appearing treatment that does not contain the test drug". At another place it claims that the placebo should be "inert". Thus, the E10 understanding of placebo is equal to the definition in 21 CFR 314.

[1]Now: International Council of Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.

3. IMPACT OF THIS DEFINITION

Can an *inactive preparation* or an *inert* one really *act*? Even from a linguistic perspective such thoughts appear to be odd. If the whole preparation is *inactive*, it simply *cannot be effective*! If differences are observed, e.g. compared with a no treatment control group, then there are only two valid conclusions:

1. The effect reflects the sum of many kinds of bias (including things like suggestive power of the physician), thus, it should be explained by the design of the trial.

2. The characteristic "inactive" should be challenged.

The bias problems might always be present, or at least, should always be anticipated. Researchers, reviewers, and readers of clinical trial reports should carefully challenge the attribute "inactive" when no treatment is given beside placebo. If there is any doubt that placebo was truly inactive, the investigation cannot reasonably by interpreted and might need to be repeated with a true placebo. As discussed below, at least commercial sponsors will have strong interests in that placebo is in fact inactive, so we can usually assume true inactiveness in most settings.

If we assume that a given placebo is in fact inactive, then it is logic that this placebo can also not have any secondary effects, *vulgo* side effects. In fact, all regulatory authorities share this fundamental view. International regulatory rules state that "events associated with placebo will usually not satisfy the criteria for an adverse drug reaction" [9, 10]. This is why authorities are not interested in "reactions" to placebos, because and as long as the placebos are in fact inactive and have been correctly manufactured and stored. If the latter, rather seldom conditions are excluded, any causality classification of an investigator becomes irrelevant, which otherwise (in case of an active product) could turn an adverse *event* into an adverse *reaction*. This means: No regulatory authority assumes that a truly inactive placebo causes any side effect. Be aware of these circumstances when interpreting articles on "nocebo" or "nocebo effects"!

Another impact of the given inactivity of placebo is that we should never anticipate that placebo *acts*! Even if we see surprising effects in a placebo group, we should start thinking that the placebo cannot be made responsible for them, instead other forms of bias, maybe including the healing power and the positive

charisma of the physician. But if you are reading articles on placebo effects or nocebo, you will make the contrary experience: Some write as if placebo itself would cause effects and if causal efficacy of placebo has been established and accepted by many researchers.

4. WHY CHOOSING PLACEBO?

Robert O'Neill, long-term chief statistician at the US FDA, strongly advised against using pseudoplacebos that appeared to become fashion in the early 1990ies. Pseudoplacebos are comparator products with suspected inferior efficacy, i.e. somewhat active and not fully inactive. He stressed that efficacy should be demonstrated by statistical hypothesis testing against an alternative treatment, where the cloud of dots after treatment A should be different from the cloud after treatment B. Then he argued that a commercial sponsor of a trial has to do a very good job to be in fact able to demonstrate such difference; with a sloppy conduct and too many confounders the delta, i.e. the difference between treatment A and B, might shrink. Hence, any commercial sponsor should have greatest interest in choosing a true placebo that is in fact inactive and not a pseudoplacebo, as with the latter the expected delta must be smaller than with placebo, increasing the required sample size or otherwise endangering the aims of the trial.

Beforehand, a "no treatment" might be a fair (a counteracting product would be unfair) and reasonable control group. But, the fact that people will usually rapidly recognise the identity of that group, such group will be prone for expectation bias. Therefore, it is usually much more reasonable to test against placebo that mimics the active product (appearance (weight, size, colour), including freedom of smell and taste or mimicking such of the active compound) and to keep all parties that could influence the data (patients, investigators, study nurses, monitors, study managers, data managers etc.) blinded for the true identity; because of historical considerations such feature is still only called "double-blind". In other words: Placebo is a key feature to control bias in a clinical trial. Blinding is another important feature that should be applied, if possible, in particular when using placebo. Nonetheless, most experts in clinical research might value randomisation (to reduce allocation bias) even on top of these bias reducing methods. All three together (randomisation, placebo, blinding) will result in a trial design that might adequately reduce and control bias, if no other fundamental errors are implemented.

All in all, placebo is used for two reasons:

- To provide the largest delta in the fairest comparison.
- To control bias.

With respect to safety assessments, placebo is important "to show lack of difference (of specified size) in evaluating a safety measurement" [6]. Again, authorities waste no time in arguing about "nocebo effects" or similar aspects.

Think of the successful mega-trials e.g. with statines (4S, Jupiter) or antidiabetics (EMPA-REG). Can you imagine that controlling of psychological effects has been considered as reason to test against placebo? I think nobody would assume psychological effects in such settings. Nevertheless, regulators required testing against placebo and companies invested many dollars to get data on a product (placebo) they will never market.

5. PLACEBO EFFECT OR BIAS?

From the above mentioned ICH guidelines, only E10 [6] briefly discusses the term "placebo effect" and states that this might be considered as "improvement in a subject resulting from thinking that he or she is taking a drug". Then the guideline stresses "but that is not its only or major benefit" of using placebo. They also considered the controlling effect of the patients' expectations as minor.

Thinking of indications where subjective variables are in the focus, such as any kind of pain or other symptoms, the patients' thinking about the test drug might be important. But there were and will be many clinical trials where subjective variables are less important or even do not play any role. Nevertheless, other forms of bias should be envisaged in all types of studies. ICH E10 [6] mentions "spontaneous change (natural history of the disease and regression to the mean), subject or investigator expectations, the effect of being in a trial, use of other therapy, and subjective elements of diagnosis or assessment".

5.1. Confounders

Even this list of potential biases or confounders might not be exhaustive. Think for instance of a rain front in a trial in allergic rhinitis that may quantitatively eliminate pollen from the atmosphere; such weather will certainly improve symptoms, but nobody would blame psychological effects for the respective

improvements in the symptoms score, which would be substantial. Even without a rain front, the burden of specific pollen will have its peak just for botanical reasons. Hence, symptoms will improve anyhow after few weeks. Other external confounders could occur in any trial, off course, in a relevant manner rather in long-term trials.

5.2. Subjective Variables and Inclusion Criteria

Clinical trials often measure “continuous variables” rather than events. Then, changes from baseline are usually calculated (yielding some intrasubject adjustment) and subsequently the treatments will be compared using these differences. However, it appears that some neglect that a difference could be influenced by the baseline reading as well as by the treatment effect at the last observation.

In fact, bias is very often set by the inclusion criteria of clinical trials, in particular if the entry criterion is equal or closely related to the primary variable. This problem is further enforced if the inclusion criterion is not objectively determined. The rationales for such

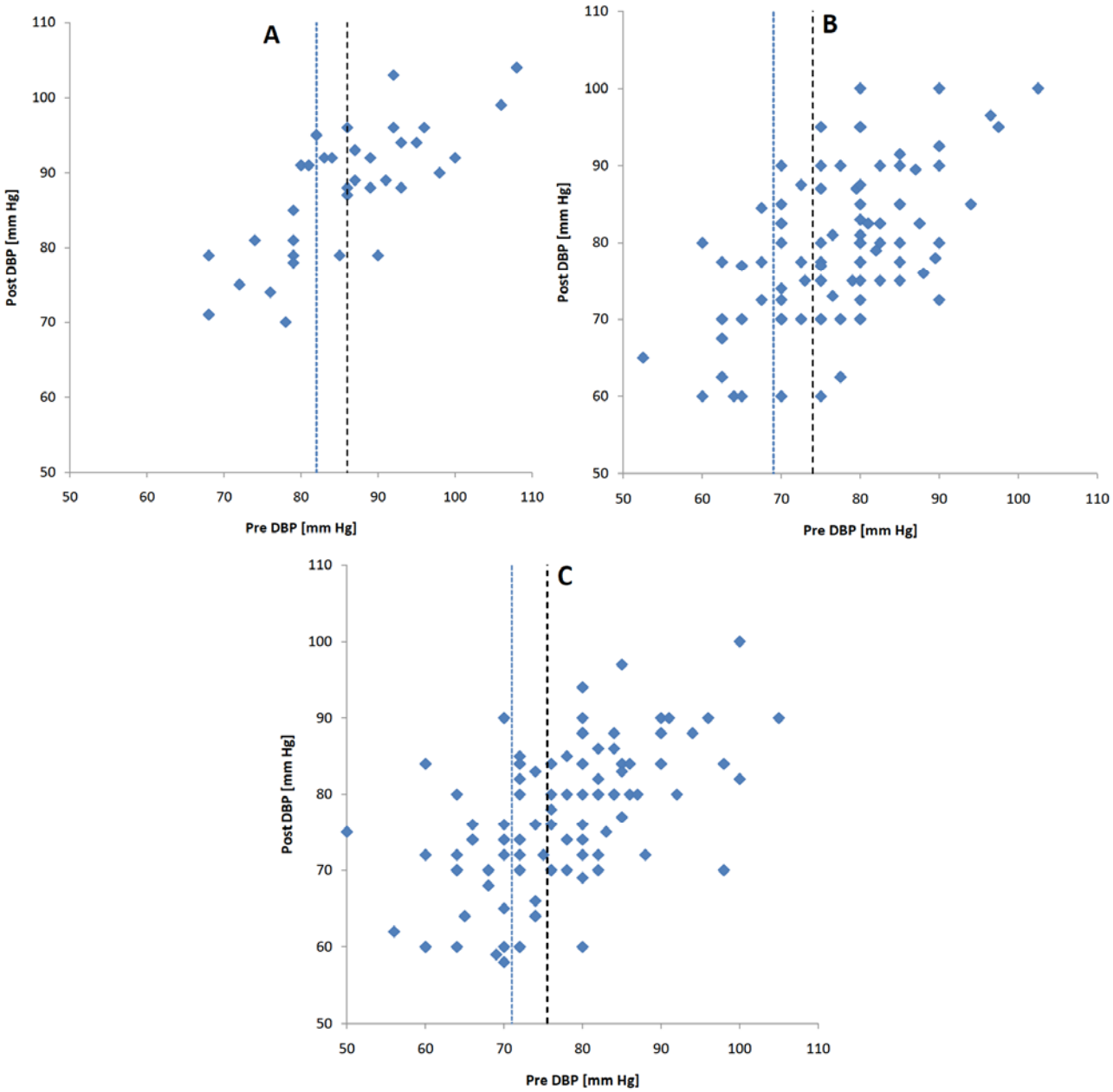

Figure 1: Effect of cutting rules on clouds of real data from 3 different trials

The variable was diastolic blood pressure (DBP) from the screening phase of controlled clinical trials. “Pre” represents the first measurement during Screening, “Post” the second. Panel **A** was from a trial in hypertension, **B** from a trial in epilepsy, and **C** from a trial in diabetic neuropathy. For statistical parameters see Table **1**.

"cutting" criteria are always, simply and true: We need to get patients with the target indication, and we should avoid treating healthy paitents.

In the 1970ies to about mid 1990ies most hypertension trials set a lower limit of diastolic blood pressure (DBP) e.g. at 90 mg Hg, that time usually measured by the physicians themselves using sphygmomanometers. There were and still are two problems related to exclusion (cutting) criteria addressing the target variable. a mathematical-statistical and an investigator problem.

The mathematical-statistical problem is illustrated by Figure **1** and Table **1** using real data of DBP measurements from the screening phases of three clinical trials in different indications. Imagine that these clouds would represent a placebo group with baseline (pre) and endpoint assessment (post) and a representative sample. Then imagine that all 6 cutting rules would exclude dots (patients) with too low baseline (pre) values from the clouds. For each panel two different cutting rules are shown, the first at about the quartile or tertile of the cloud, the second close to the mean of the original data set. Each cutting rule decreases the mean post-pre-value compared with the original cloud, although in fact no active treatment can be made responsible. The closer the cut to the mean the stronger; reaching "effect sizes" of a -2 to -3 mm Hg (Table **1**). Cutting the clouds in the contrary sense (e.g. in the Panel C all data pairs with pre-levels equal or above 80 mm Hg) yields an increase in the post-pre-change of 2.4 mm Hg (data not shown).

The explanation for this phenomenon is that we can only exclude based on the first values in the time course, but these values show the similar fluctuations as the later assessed values. By excluding the low values on the first assessment, we do not only exclude the *truely low* values, but also the *by chance low* values, and we leave in the *truely high* values (as intended) and the *by chance high* values. Such selection causes a bias. Hence, each variable-based exclusion criterion has impact on the mean changes over time of that variable. The extend of the impact depends on the distance between the criterion and the true mean, the direction of the impact depends on the direction of the cutting rule.

Usually, trialist have little chance in estimating such effects from their trials, as the "true" mean of the given population might remain unknown, even more, if a respective exclusion criterion was set. In some indications, in particular, if high symptoms scores are required (e.g. in pain, allergic rhinitis, neuropathy, depression etc.) the problem may become substantial as the cutting criterion will often be very close to the "true" mean (e.g. at least a "moderate" when the score is none, mild, moderate, and severe). In simplistic words: In such situations the placebo group has little chance to show anything else than improvement – just be design and mathematics.

The other and often more important effect is rather related to investigators: Did they not have an interest

Table 1: Effect of Cutting Rules: Statistical Parameters Pertinent to Figure 1

		Pre	Post	Post – pre
Panel A	**All Data**			
Tiral in	N	36	36	36
Hypertension	Mean	86.19	87.94	1.75
	SD	9.43	8.62	6.35
	Cut at 82 (pre)			
	N	24	24	24
	Mean	91.25	92.65	0.88
	SD	6.73	6.00	6.54
			Effect of cutting:	*-0.88*
	Cut at 89.5 (pre)			
	N	20	20	20
	Mean	92.80	92.65	-0.15
	SD	6.27	5.82	5.86
			Effect of cutting:	*-1.90*
Panel B	**All Data**			
Trial in	N	100	100	100
epilpsy	Mean	76,96	78,4	1,44
	SD	8.88	9.76	8.52
	Cut at 69 (pre)			
	N	85	85	85
	Mean	79.41	79.92	0.51
	SD	7.08	9.25	8.36
			Effect of cutting:	*-0.92*
	Cut at 74 (pre)			
	N	67	67	67
	Mean	81.78	81.30	-0.48
	SD	6.05	9.01	8.13
			Effect of cutting:	*-1.92*
Panel C	**All Data**			
Trial in	N	100	100	100
diabetic	Mean	77.01	76.71	-0,3
neuropathy	SD	10.64	9.46	9.56
	Cut at 71(pre)			
	N	70	70	70
	Mean	82,1	79,37	-2,73
	SD	8,05	8,64	8,6
			Effect of cutting:	*-2.43*
	Cut at 76 (pre)			
	N	55	55	55
	Mean	84.62	81.16	-3.45
	SD	7.24	7.77	8.43
			Effect of cutting:	*-3.15*

Mean, SD, and effect refer to DBP in mm Hg. *Effect of cutting* represents the difference of the mean after cutting minus the mean of the all data.

"to read" a level above, e.g., 90 or 95 mm Hg at the sphygmomanometer (now also called: "clinical blood pressure" [11]), given that they had informed the patients about the trial and done all the baseline assessments? That they hoped to be able to recruit an(other) patient?

Both effects, the mathematical statistical and the investigator-related effect, were and, in similar settings, are still working together resulting in a "placebo effect", e.g. in an ACE-inhibitor trial [12] of about 4 mm HG in the placebo group when relying on such supine "clinical blood pressure", as compared with about 8 mm HG after a, now recommended, 20 mg dose of the tested ACE inhibitor, yielding a difference of about 4 mm Hg between placebo and active. Later, a Cochrane review [13] estimated the effect size of ACE-inhibitors as about 5 mm Hg on diastolic blood pressure, without evidence for substantial differences between different active compounds. Meanwhile, treatment guidelines for hypertension request assessment of the less biased machine-assessed "ambulatory blood pressure" in addition to the "clinical blood pressure" [11], namely because such ambulatory blood pressure eliminated at least the investigator-related "placebo effects", at in hypertension trials if adequately used [14]. It may be that psychology (physicians' expectations, care and devotion and patients' hopes) might have contributed to the overall effect of ACE-inhibitors on "clinical blood pressure" of 8 to 9 mm Hg, but as the example shows, we do not need such explanations. Why should we then speculate about the wonderful effects of placebo or medical care?

In many indications, researchers are not in the comfortable position to be able to switch simply to an automatic system. Particularly, if symptoms are in the focus, important bias because of "baseline reading" should be envisaged, in these cases caused by investigators and possibly also patients. This problem might be enforced, if a symptoms score is an inclusion criterion, a dilemma that can often not be resolved.

Some feeling of the size of the effect of inclusion criteria even on objective variables might be derived from the very large Jupiter trial [15]. It investigated the long-term effects (mortality etc.) of rosuvastatin vs. placebo in 17802 patients. The main inclusion criterion, in fact the idea of this study, was a CRP of at least 2.0 mg/l, the other important inclusion criterion was an LDL(-cholesterol) level of less than 130 mg/l. If the above considerations were relevant, we should expect a decrease in CRP and an increase in LDL. In fact, after 12 months (when most patients were still in the trial and adherence might have been better than later on) the placebo group showed a decrease in CRP, with means falling from 6.9 to 6.0 mg/dl [16]. Vice versa, the lipid levels showed increases in the means: LDL from 105.6 to 109.1 mg/dl, the HDL from 51.3 to 52.2, and the total cholesterol from 183 to 189 mg/dl [16]. Due to the extremely large sample sizes all these changes were "highly significant", despite the fact that most would classify them as marginal or clinically irrelevant. As often, triglyceride levels showed high variability preventing reasonable conclusions.

5.3. Other Confounders

Many trials, in particular those testing analgesics, allow for rescue medication such as paracetamol (acetaminophen). Clearly, its use should and will be documented in the trial. But, up to now, nobody adjusts the actual pain scores by the use of rescue medication. So, the average improvement in the pain score will also reflect the effect of the rescue medication, both in the placebo and the active group. Hence, be very careful before attributing pain improvements to psychology!

Placebo-controlled trials are usually superiority trials. With this question (active is better than placebo) we have to analyse the data according to the intention to treat principle. Imagine a patient in such trial is noncompliant because he thinks the drug is ineffective: Many investigators would discontinue the patient, unfortunately sometimes without asking for the reason of the noncompliance. The problem is minor, if a patient discontinues "due to lack of efficacy". If such discontinuations were more frequent on placebo than on active, then we would have captured evidence for efficacy, although maybe not in the prospectively determined primary variable. In reality, such patients often discontinue due to "noncompliance" alone, or they will be retained although one should assume that they have sought and used other effective treatments. Often it is asked to categorise the reason for drop-out into primary or secondary reasons, although more than one reason may apply equally. Then, information can get lost. Have such circumstances always be considered when claiming superb effects of placebo?

A given clinical trial protocol is hopefully excellent, but not always. And any protocol must be interpreted. In an ideal world, everything should be clear and under control. In the real world investigators, monitors, and contract research organisations have to interpret the rules, sometimes with questionable interpretations, causing bias; another one that is controlled by placebo.

5.4. The Issue with the Blinding

As a rule, we should be more critical with blinding issues. Researcher and pharmacies may have done an excellent job in this respect, but a patient who is experienced with, e.g., allergic rhinitis and its treatment might be able to distinguish placebo from active. Not all patients, but some. In the past we only rarely got data on this. If such an experienced patient assumes receiving placebo because of the lack of rapid relief of symptoms: How willing to continue administering the seemingly inactive test drug he will be? Can we exclude that he will take medication of which he knows that it works? This certainly will occur. Unfortunately, not all patients will tell this and sometimes such telling might not be encouraged by protocol or patient information. Anyhow, from the published material we usually do not get such very specific information.

Even more problematic is the situation in many dermatological trials. With modern drugs such as imiquimod or ingenol mebutate local inflammation is provoked that may heal the cancer or a precancerous condition. The physicians know this and, hence, the patients should be informed respectively. Will, however, every patient enrolled in the trial be fully compliant if recognising that nothing happens, although inflammation is expected and the most reasonable explanation is: randomised to placebo (= vehicle)? Remember: These patients have cancerous conditions!

All in all I would stress: The effects observed in the placebo group reflect the sum of many kinds of bias, some we can expect prospectively, and some we know little about and we can often not imagine during planning. Psychological bias (suggestion, care, devotion, expectation, hope) might be present, but at least in the examples above, these aspects are not necessary to explain the “placebo effects”.

6. THE TRUE EFFECT SIZE

As explained above, changes from baseline without a control group might be biased, in the placebo group as well as in the active group. Therefore, the hypothesis test will always compare two treatments, while an isolated change from baseline might be hypothesis generating, not more. The absolute effect size is often not in the focus of a controlled clinical trial. If, however, you read articles on placebo effects, the authors argue often with changes from baseline in the placebo group. Such changes might, if any, became interpretable in 3-arm designs with a “no treatment contol”, and even then, we should assume that the data are biased because patients recognise the difference between no treatment and treatment with blinded study medication; and usually also the investigators.

Considering the phenomena discussed above, the true effect size in the usual 2-arm design can only be determined if all following conditions apply:

- It is unlikely that bias is exerted by selection criteria.
- It is unlikely that bias is exerted by other design aspects.
- The variable is objectively assessed.

Vice versa, if one of these conditions cannot be excluded, we should envisage biased measurements for all treatments, including placebo.

There are many variables in modern clinical research, where bias is rather unlikely. This applies often for the so-called “hard endpoints”. However, already deaths due to specific causes might often carry the risk of subjectivity, namely due to the medical interpretation of the individual cases. This means, we should not simplify and exculpate all “hard endpoints” from concerns of bias. Death due to any cause (i.e. all-cause mortality) might have very low, if any, risk of bias. Fortunately, this variable is directly correlated with survival. And is this not the variable of ultimate interest for the patient in many (not all!) conditions?

We should be very careful with predicting from placebo controlled clinical trials that Wonderdrug A causes x.y percent improvement in symptoms. If the difference to placebo was statistically significant, we can usually only say that it might be x.y score points or maybe x.y percent better than placebo, i.e. it is effective.

7. BIAS AND GCP

Many may think that pharmaceutical industry has most interest in biasing data. Their financial interest is obvious and had led to worldwide implementation of Good Clinical Practice by ICH E6 after 1995 [3], while the basic rules had been implemented much earlier in the USA. Certainly, these GCP rules, in particular the verification of the data with the source by monitors, audits of all processes of a clinical trial from source data over data management to data in tables, and finally the check of all these activities by inspections and re-analysis of the data by regulatory authorities made it much more difficult, if not impossible, to invent,

fabricate, or falsify data. In addition, exaggerations in publications are usually tackled by FDA and other authorities. Maybe not everything is fine nowadays, but the situation has been tremendously improved since at least the 1970ies in company-sponsored clinical research.

While any pharmaceutical manufacturer must guarantee adherence to E6 when submitting a dossier to authorities, academics usually do not have to make such submissions and are therefore (fourth paragraph of E6 [3]) out of this game. Even more disturbing is that the current CONSORT statement [17] in no way encourages mentioning such measures. It appears that we should assume that academic investigators are per se independent and have no conflict of interest, maybe because they are all and always fully interested in the truth, and nothing but the truth. There is, however, much contradicting evidence.

A fine example suggesting the contrary for academic clinical trials provides the DAMASCENE meta-analysis [18]. Those authors wanted to calculate overall estimates for the beneficial effects of stem cell therapy for heart failure and similar conditions. Instead, they found many discrepancies within and between reports of the same trial. Then they plotted the number of discrepancies against the effect size: The more discrepancies, the stronger the effect, with almost no effect when no or few discrepancies were found. The points for me are: None of these trials were sponsored by pharmaceutical companies, hence, monitoring, auditing, and concise quality control were no "must haves" for most of these trials. Investigators have fundamental interests in publishing sound articles. At least statistical evidence supports the assumption that sometimes data are a bit stretched to yield the intended outcome.

With respect to general biomedical research (nonclinical and clinical), John Ioannidis published a lot of evidence that really many published research findings are false [28]. When industry repeated academic research, the reproducibility rates were very low, e.g. 11% when done by Amgen and 25% when done by Bayer [29]. It should also be reminded to a very broad examination of psychological studies published recently [30], indicating that the results of less than half of 100 original studies could be reproduced.

These aspects should be kept in mind when discussing placebo research. If the study did not have monitoring, auditing, and concise quality control, this should be taken as a warning signal.

8. PLACEBO RESEARCH USING A 3-ARM DESIGN

The only reasonable design to investigate "placebo effects" is a three arm design, with one most likely active group, the matching placebo group, and a no treatment control [27]. Even then, an important bias cannot be excluded, namely due to knowledge of and within the no treatment group.

Hróbjartsson and Gøtzsche provided excellent meta-analyses on such trials [19-22]. In the 2010 update [22] they analysed 202 trials and found no important clinical effects in general. As the theoretical considerations predict, subjective variables, in particular pain and nausea, showed the "strongest effects" although it was "difficult to distinguish patient-reported effects of placebo from biased reporting". The effect on pain varied, even among trials with low risk of bias, from negligible to clinically important, indicating that design aspects might be of overwhelming importance.

My interpretation is: High-quality research found no firm evidence for a true placebo effect. This is in line with what we should expect simply from the definition of placebo.

9. OTHER EVIDENCE FOR PLACEBO EFFECTS?

Despite the theoretical background and despite these excellent meta-analyses, the belief in placebo effects seems to be stronger. There is still a plethora of articles on placebo, and most of these still refers to studies that claimed these or those effects, sometimes even objective effects (as e.g. outlined in the English Wikipedia article on placebo). However, most of these isolated research articles appear to be flawed:

- Often the authors anticipate the existence of a "placebo effect" in a given indication. Already such expectation should not be present in independent research. It indicates lack of scientific neutrality and suggests a bias for a certain outcome.
- Maybe related to the former concern, many trials were designed with only 2 arms.
- Maybe also related to the first point, trials aiming to demonstrate objective effects (e.g. on endorphins etc.) often did not attempt to measure the symptomatic counterparts, i.e.

testing whether the basic assumption (placebo is effective in pain relief) would be true in their model.

- Many trials used cross-over designs with important flaws. For single dose administrations with 2 treatments such designs might work, however, with 3 treatments or even multiple doses (i.e. several days of treatment) the risk for drop-outs increases, rendering the interpretation of the trial at least problematic. In fact, many of these trials had drop-outs, and usually too sparse details on drop-outs or on individual data were presented. In addition many papers did not provide anything on sequence or carry-over effects, issues that themselves could render the data not interpretable.

- Virtually none of the mechanistic trials was sponsored by a pharmaceutical company. Hence, almost all these trials are lacking monitoring, auditing, and concise quality control, not even speaking about FDA inspections or control of such publications.

Only few trials remain that deserve further notice.

Often cited is a 3-arm trial in irritable bowel disease [23]. Because placebo appeared to be significantly effective, the authors even claimed that it could be used "without deception" in this indication. However, critics on the journal's homepage stressed that they used cellulose as placebo. Meanwhile, cellulose can no more be considered ineffective in this indication [24].

Another, beforehand surprising outcome was reported from a study of testing cough treatment in babies and young children with agave syrup [25]. The authors found that placebo and a bit more the agave syrup were significantly different from no treatment, i.e. already the placebo alone appeared to be effective. The rationale for using the agave syrup was the high content of sugar, and that honey (otherwise classified as possibly effective [26]) would not be optimal for children. The authors stated that "natural grape-flavored water with caramel color" was used as placebo. However, can "natural grape-flavored water" ever be produced without glucose? At least I am challenging the attribute "inactive" here. Most likely this placebo problem might have been anticipated and might have been the reason for testing this question in a 3-arm design.

Finally a note on endorphin expression claimed by some to occur after placebo. Even if this would be true and attributable to placebo or, more correctly, to the stories the investigators tell the research subjects or patients: Is this of any help to explain the "placebo effect" in indications other than pain?

10. PLACEBO AND NOCEBO

Wikipedia defines "nocebo" as "an inert substance or form of therapy that creates harmful effects in a patient". End of October 2015, the English Wikipedia article on placebo had about 5900 words (without the 175 references). At the same time, the article on "nocebo" had about 2159 words (without the 22 references). At the time, PubMed indicates about 185 000 hits for "placebo" and 68 000 hits for "placebo controlled", compared with only 331 hits for nocebo and not any hit for "nocebo controlled". These figures alone might challenge the scientific relevance of the term nocebo. In fact, this term is a concept rather than a true research term. None of the before mentioned international guidelines even contain the word nocebo. Finally, Pub Med lists 61 hits for "nocebo" if limited to publication after 2014, indicating that this term becomes something like the *denier cri*.

Going more into depth, the things get worse. English Wikipedia refers e.g. to a meta-analysis of adverse events in trials in Parkinson's disease [31]. The authors fundamentally err in that they assume that patients truly discontinue due to "placebo intolerance" as written in the abstract. If this would be true, all regulatory authorities would be false (as outlined above). Maybe that some discontinuations had been attributed to the test product and that it later turned out that this was in fact placebo. But in truly blinded trials, patients and investigators must do a causality assessment under blind conditions. Are they not allowed to err here? In fact, the authors analysed any adverse *events* and discontinuations due to adverse *events*, this means including e.g. drop-outs due to myocardial infarction, surgery, cancer, or death; events in which most investigators would withdraw a patient from a trial and in which only very few would suspect a *reaction* in an individual case. Hence, in no way these figures reflect "placebo intolerance" or "patient's negative expectations" as suggested in the abstract, but more or less what happens in real life. Moreover, the high inter-trial variability and the possible basic treatment with other drugs were disregarded. The concerns are similar in another meta-analysis published by this group [32].

A meta-analysis of depression trials from another group appears to be more honest [33]. These authors wrote that "TEAEs[2] were very common among placebo-treated clinical trial participants. Unexpectedly, there was no evidence to associate TEAEs with adverse clinical outcomes, nor were the conditioning or expectancy hypotheses supported by these data." Instead, I am wondering why this outcome surprised the authors and why they are still assuming a nocebo effect and requesting further research.

Some of those claiming nocebo effects refer to a study of the effects of electromagnetic fields (EMF) with mobile phones [34], although these authors did not use this term themselves. In fact, this study might be interpreted in this or that way with respect to EMF (a major problem were the unbalanced drop-outs and lack of respective details), but any conclusion about placebo or nocebo effects appear to be unjustified. A recent study confirmed that EMF cannot be identified by human beings [35], maybe not so surprising as we do not have a respective sense. Even the "effects" observed in the "sensitive" subject indicate only their individual prior perception and not an influence of investigators or the study situation. And again: Knowing the type of exposure caused strong effects. As this outcome is derived from unblinded observation, we should not blame placebo or nocebo for this.

11. CAN PLACEBO BE PRESCRIBED?

No and yes!

No, because no drug legislation allows marketing of placebos. The rationales are: What should the physician tell the patient? Something else than the truth? What if the patient reads "placebo"? No way!

A conditioned yes, because in most countries physicians may prescribe homoeopathics. Although this theme usually causes endless discussions as well, most might agree that with "high potencies" no effect should be expected. Moreover, those who gladly prescribe homoeopathics would argue: I do not know whether it works, but at least it causes no harms.

Another conditioned yes, because sometimes there are in fact possibilities to prescribe placebo. This could be a basic cream (vehicle) of a cortisone cream or just a saline infusion. Finally, the physician could order individual manufacturing by a pharmacy of, e.g., lactose capsules. However, in these examples the patient might recognise the nature of the product.

12. CONCLUDING REMARKS

The perception that medical care as such and personal attention or devotion could heal or at least help patients is certainly very attractive to physicians and caregivers. Maybe this is the ultimate reason for so many to belief in the importance of the "placebo effect" or vice versa "nocebo effects". There are many anecdotes on phantastic healing without having administered an established medical intervention, some may call them "wonders". Even most atheists would accept that faith can move mountains. Maybe that medical care and personal devotion and the patient's beliefs contribute a lot in individual cases. Yet, adequately derived and interpreted data from placebo controlled trials provide, if any, little evidence that psychological factors are important on average. Vice versa, such factors are only the minority among a plethora of other, often more powerful sorts of bias. In many conditions and research questions, psychological factors might have no relevance at all, at least as long as blinding can be maintained to everybody involved in the research until database lock. Thinking that mainly psychological effects need to be controlled by placebo appears to be narrow-minded.

We should adhere to the given definition, namely that placebo is an inactive preparation, and never assume the contrary.

Physicians and patients should be aware of what is placebo for and of what can be derived from placebo controlled trials. Placebo-controlled trials are performed to confirm a difference to an active drug, including to provide a reliable estimate for the difference between active and placebo. Randomisation, placebo, and blinding are essential to reduce and control bias. The absolute point estimates from such trials, however, can often not be easily interpreted.

CONFLICT OF INTEREST

None.

FUNDING

No external funding.

REFERENCES

[1] McDonald CJ, Mazzuca SA, McCabe GP Jr. How much of the placebo 'effect' is really statistical regression? Stat Med 1983: 2(4): 417-27. http://dx.doi.org/10.1002/sim.4780020401

[2]Treatment emergent adverse events.

[2] Boutron I, Estellat C, Guittet L, *et al*. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. PLoS Med 2006; 3(10): e425. http://dx.doi.org/10.1371/journal.pmed.0030425

[3] ICH E6: Guideline for Good Clinical Practice. May 1996, as adopted including including post step errata, status July 2002. Available from: www.ich.org

[4] ICH E8: General Considerations for Clinical Trials. July 1997. Available from: www.ich.org

[5] ICH E9: Statistical Principles for Clinical Trials. February 1998. Available from: www.ich.org

[6] ICH E10: Choice of Control Group in Clinical Trials. July 2000. Available from: www.ich.org

[7] European Commission: Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. Published in the Official Journal L – 121, 01 May 2001, p. 34-44.

[8] European Commission Regulation (EU) No 536/2014* of the European parliament and of the council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC, Published in the Official Journal L 158/1, 27 May 2014, p. 1-76.

[9] ICH E2A: Clinical safety data management: definitions and standards for expedited reporting (step 4: November 1994, step 5 (EU): June 1995).

[10] European Commission: Communication form the Commission – Detailed guidance on the collection, verification and presentation of adverse reaction reports arising from clinical trials on medicinal products for human use ('CT-3'). 2011/C 172/01, Official Journal 11 June 2011

[11] National Clinical Guideline Centre. Hypertension. Clinical management of primary hypertension in adults. London (UK): National Institute for Health and Clinical Excellence (NICE); 2011 Aug. 36 p. (Clinical guideline; no. 127). Available from: http://www.nice.org.uk

[12] Gomez HJ, Cirillo VJ, Sromovsky JA, *et al*. Lisinopril dose-response relationship in essential hypertension. Br J Clin Pharmacol 1989; 28(4): 415-20. http://dx.doi.org/10.1111/j.1365-2125.1989.tb03521.x

[13] Heran BS, Wong MM, Heran IK, Wright JM. Blood pressure lowering efficacy of angiotensin converting enzyme (ACE) inhibitors for primary hypertension. Cochrane Database Syst Rev 2008; (4): CD003823. http://dx.doi.org/10.1002/14651858.CD003823.pub2

[14] Dupont AG, Van der Niepen P, Six RO. Placebo does not lower ambulatory blood pressure. Br J Clin Pharmacol 1987; 24(1): 106-9. http://dx.doi.org/10.1111/j.1365-2125.1987.tb03145.x

[15] Ridker PM, Danielson E, Fonseca FA, *et al*. JUPITER Study Group. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 2008; 359(21): 2195-207. http://dx.doi.org/10.1056/NEJMoa0807646

[16] Public Assessment Report of the Medicines Evaluation Board in the Netherlands. Crestor 5/10/20/40, 5/10/20/40 mg film-coated tablets, Astra Zeneca B.V., The Netherlands. 3 February 2011. Therein page 64. Available from: http://www.cbg-meb.nl/

[17] Schulz KF, Altman DG, Moher D. CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010; 152(11): 726-32. http://dx.doi.org/10.7326/0003-4819-152-11-201006010-00232

[18] Nowbar AN, Mielewczik M, Karavassilis M, *et al*. DAMASCENE writing group. Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis. BMJ 2014; 348: g2688. http://dx.doi.org/10.1136/bmj.g2688

[19] Hróbjartsson A, Gøtzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med 2001; 344(21): 1594-602. http://dx.doi.org/10.1056/NEJM200105243442106

[20] Hróbjartsson A, Gøtzsche PC. Placebo treatment versus no treatment. Cochrane Database Syst Rev 2003; (1): CD003974. http://dx.doi.org/10.1002/14651858.cd003974.pub2

[21] Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. Cochrane Database Syst Rev 2004; (3): CD003974.

[22] Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. Cochrane Database Syst Rev 2010; (1): CD003974. http://dx.doi.org/10.1002/14651858.CD003974.pub3

[23] Kaptchuk TJ, Friedlander E, Kelley JM, *et al*. Placebos without deception: a randomized controlled trial in irritable bowel syndrome. PLoS One 2010; 5(12): e15591. http://dx.doi.org/10.1371/journal.pone.0015591

[24] Moayyedi P, Quigley EM, Lacy BE, Lembo AJ, Saito YA, Schiller LR, Soffer EE, Spiegel BM, Ford AC. The effect of fiber supplementation on irritable bowel syndrome: a systematic review and meta-analysis. Am J Gastroenterol 2014; 109(9): 1367-74. http://dx.doi.org/10.1038/ajg.2014.195

[25] Paul IM, Beiler JS, Vallati JR, Duda LM, King TS. Placebo effect in the treatment of acute cough in infants and toddlers: a randomized clinical trial. JAMA Pediatr 2014; 168(12): 1107-13. http://dx.doi.org/10.1001/jamapediatrics.2014.1609

[26] Oduwole O, Meremikwu MM, Oyo-Ita A, Udoh EE. Honey for acute cough in children. Cochrane Database Syst Rev 2014; 12: CD007094. http://dx.doi.org/10.1002/14651858.CD007094.pub4

[27] Senn S. Placebo misconceptions. Am J Bioeth 2009; 9(9): 53-4. http://dx.doi.org/10.1080/15265160903098283

[28] Ioannidis JP. Why most published research findings are false. PLoS Med 2005; 2(8): e124. http://dx.doi.org/10.1371/journal.pmed.0020124

[29] Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature 2012; 483(7391): 531-533. http://dx.doi.org/10.1038/483531a

[30] Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science 2015; 349(6251): aac4716. http://dx.doi.org/10.1126/science.aac4716

[31] Stathis P, Smpiliris M, Konitsiotis S, Mitsikostas DD. Nocebo as a potential confounding factor in clinical trials for Parkinson's disease treatment: a meta-analysis. Eur J Neurol 2013; 20(3): 527-33. http://dx.doi.org/10.1111/ene.12014

[32] Mitsikostas DD, Mantonakis L, Chalarakis N. Nocebo in clinical trials for depression: a meta-analysis. Psychiatry Res 2014; 215(1): 82-6. http://dx.doi.org/10.1016/j.psychres.2013.10.019

[33] Dodd S, Schacht A, Kelin K, *et al*. Nocebo effects in the treatment of major depression: results from an individual study participant-level meta-analysis of the placebo arm of duloxetine clinical trials. J Clin Psychiatry 2015; 76(6): 702-11. http://dx.doi.org/10.4088/JCP.13r08858

[34] Eltiti S, Wallace D, Ridgewell A, *et al*. Does short-term exposure to mobile phone base station signals increase symptoms in individuals who report sensitivity to

electromagnetic fields? A double-blind randomized provocation study. Environ Health Perspect 2007; 115(11): 1603-8.
http://dx.doi.org/10.1289/ehp.10286

[35] Malek F, Rani KA, Rahim HA, Omar MH. Effect of Short-Term Mobile Phone Base Station Exposure on Cognitive Performance, Body Temperature. Heart Rate and Blood Pressure of Malaysians Sci Rep 2015; 5: 13206.

Received on 31-10-2015 Accepted on 29-02-2016 Published on 27-04-2016

DOI: http://dx.doi.org/10.6000/1927-5951.2016.06.02.2