BLINDING AND THE NON-INTERFERENCE ASSUMPTION
IN MEDICAL AND SOCIAL TRIALS

DAVID TEIRA

(UNED, dteira@fsof.uned.es)

1. INTRODUCTION

Randomized trials are experiments in which in which we test the comparative efficacy of two different treatments measuring their effects on two groups of people. Each intervention is administered to a group of participants, chosen at random; the effects are then measured and compared, testing the statistical significance of the difference. The design of the experiment has been used in medicine for more than six decades now in order to test the safety and efficacy of an innovative therapy as compared to a given standard (or a placebo). Field experiments in the social sciences apply this experimental design to test policy interventions aimed at a given goal. The assessment of public policy programs through large-scale field experiments is already several decades old (the 1968 New Jersey negative income tax experiment is considered a pioneering example). Around 200 randomized field trials were run in the United States between 1960 and 1999, with more or less convincing results. In the last decade, there has been an explosion of interest in field experiments, mainly among development economists. Several programs for improving health or education, different microfinance and governance schemes have been tested in a number of developing countries. Many field experimentalists in economics expect their studies to test tools to fight poverty at least with the same degree of reliability than randomized clinical trials test drugs in medicine.

I want to discuss here the so-called *non-interference assumption* (NIA) that grounds causal inference in trials in both medicine and the social sciences[1]. This assumption states that for each participant in the experiment, the value of the potential outcome depends only upon whether or not s/he gets the treatment. In other words, the potential outcomes that would arise if a subject were affected by the treatment of other

---

[1] The NIA is often known as the Stable Unit Treatment Value Assumption (SUTVA). I follow here Gerber and Green's terminology.

subjects are declared negligible. Hence, the causal interpretation of the treatment effects is not affected by the overall pattern of actual or assigned treatments generated by randomization (Gerber and Green 2012, pp. 43-44).

The NIA is quite plausible in the context of medical trials, where the therapy received by a patient should not disturb the effect of the alternative treatment: if I am taking pill A, there is no physical mechanism by which it can have an effect on your intake of pill B. This interference can only take place through the expectations of the patients about each treatment: if I think that the experimental treatment is better than the standard alternative and I believe I am receiving this latter, these expectations may impinge on the outcome[2]. There is wide evidence about such *placebo effects* and in order to prevent them, clinical trials are *double blinded*, if the therapy allows it: the treatments are masked so that participants remain ignorant of which on they are receiving. In other words, *the masking of treatments constitutes the methodological warrant of the NIA*. We still know little about the effect of the participants' expectations in field experiments in the social sciences. Drawing on the knowledge accumulated in medical trials, on the one hand, and in laboratory experiments in economics, in this paper I want to examine to what extent blinding provides a warrant of the NIA in field trials. In the rest of this introductory section let me motivate my claim more precisely.

In a minimal definition, a *double blind* trial requires participants, investigators and outcome assessors to be unaware of the patient's assigned treatment throughout the conduct of the trial (Fergusson et al. 2004, p. 1)[3]. There is evidence about medical trials showing that blinded patients may report symptoms differently from unblinded patients; they also have different thresholds for leaving a trial or seeking additional treatments outside the research protocol[4]. In order to control the patients' expectations about treatment, those in the control group receive a placebo when no effective treatment is available. Depending on the type of intervention under study, placebos may be presented as, e.g., capsules, sham devices or any other form simulating the administration of the experimental treatment. A participant is considered blinded if the allocation and the treatment are properly concealed, at least at the beginning of the trial

---

[2] We still lack a well-articulated theory about the psychological mechanisms involved in the placebo effect, so we use the folk-psychological term of *expectations* to account for it (see, e.g., Evans 2004).

[3] For a discussion of the concept, see Howick 2011, pp. 66-68.

[4] See, e.g., Schulz et al. 1995, 1996, 2002 and Gluud 2006.

(Schulz et al. 2002): the allocation should be concealed in order to avoid differences in enrollment or early withdrawals (*selection bias*); treatment is concealed in order to prevent biases originating in the interpretation of the clinical outcome (*ascertainment bias*) and subsequent withdrawals.

When entering a clinical trial, the participants should sign an informed consent form in which they learn about the expected effects of the experimental treatment and about the possibility of receiving a placebo. Following Colagiuri (2010), we may say that this information gives rise to *expectations* of two types: about the effects of the experimental treatment, on the one hand, and about the effects of the perceived treatment, on the other. We know what may happen if we get the experimental treatment, but we do not know if we are getting it. Depending on our guesses on the actually assigned treatment, the "physical" and the "imaginary" effects will appear in various combinations in each arm of the trial: e.g., some patients may think that they are getting a placebo when in fact they are receiving the true intervention, etc. The aim of a randomized allocation of concealed treatments is to distribute evenly these expectations across the arms of the trial in order to prevent them from having an effect on the outcome.

In this paper, I am going to present and defend the following claims. First, if the participants are not indifferent regarding treatments, we need to implement a blinding device in every trial in order to ground the NIA. But we cannot take its efficacy for granted: we need to test that the blinding actually controlled for the expectations of the participants and no malign unmasking spoiled the NIA. Precisely because this test is necessary, we can only blind the participants up to a certain point: we cannot *deceive* them. There is evidence showing that if they suspect they are being deceived, they will deviate from the trial protocol, flawing the outcome.

In section 2, I will present a crucial distinction for my case. The participants have two types of expectations that trialists should control for. They have expectations about the *outcome* of an experiment (e.g., the treatment they may receive): blinding controls for them in order to sustain the NIA. But the participants also have expectations about the experimental *procedure*: they may care about being randomized or blinded. Expectations of this latter type give rise to the so-called *experimenter demand effects* (sometimes, less precisely, *Hawthorne effects*): the participants are second-guessing the intentions of the experimenter in a way that alters their behavior. They do not act as

they would outside the experimental setting, compromising the validity of its results. In sections 3 and 4, I will discuss how these procedural expectations set limits on the amount of blinding we should impose on our experimental subjects: we need to test the success of blinding (in order to check the NIA) and we can only do this with questionnaires, so we need our participants to answer sincerely. If they suspect they are being deceived, they may lie. Hence, field trialists should resist the temptation to deceive their patients, as they sometimes suggest.

2. THE EXPECTATIONS OF THE PARTICIPANTS IN A TRIAL

The participants in comparative trials usually have expectations about the *outcome* of the experiment and it may be dramatically different depending on the treatment they are receiving. The case of drug trials in developing countries illustrates this point. These are countries in which there are fewer opportunities for the majority of the population to access decent treatments, even for standard conditions. Clinical trials provide a unique occasion for them to gain access to new drugs, even in the testing stage, but the participants find themselves in a position of inferiority regarding the pharmaceutical companies when it comes to negotiate the details of the trial. A case in point that brought international attention to the problem was a 1997 study of maternal-to-child HIV transmission conducted in Africa: the contrasting drug was a placebo instead of the best available treatment, as it should have been according the developed world bioethical standards (Petryna 2009, pp. 33-37). But most patients did not expect any care whatsoever outside the trial and they were willing to participate in the trial and take their chances with a placebo.

But if their expectations about the treatment outcome are higher, they may not accept placebos so easily. The early anti-AIDS AZT trials conducted in the United States provide a good instance: as Epstein 1996 shows, many of the potential participants had been involved in the gay liberation movement of the 1970s and they used this network to organize themselves and make their claims about the conduct of trials. Since they did not believe in the hypothesized equivalence between AZT and placebo, many trial participants resorted to all sort of tricks to make sure they had the former (exchanging pills, taking them to labs to test whether they had an active principle, etc.). In sum, the patients distorted the research protocol to a point that made the experiment unfeasible.

Even if the two treatments are in principle equivalent, in economic field trials, people may behave differently if they realize in which group they are depending on their taste for each treatment: we can easily imagine a situation in which the participants act in a way that secures either the success or failure of a treatment, depending on their own interests. If we want to increase the number of kids that go to school in a deprived area, is it better to pay a compensatory sum to parents to give them direct incentives to send them there or shall we pay for the kids' meals at school? Even if the latter option may turn out to be more effective, parents may be too tempted by the prospect of receiving cash to acknowledge it. After all, we are testing social policies about which people may have opposite views not very sensitive to empirical evidence.

If the outcome of a trial, be it a drug or a social policy, is important enough for the participants we cannot expect them to comply with a randomized protocol, unless they believe the treatments in both arms are equivalent. In this respect, blinding is a means of protocol enforcement: if the participants do not know the treatment they are receiving, their only chance to get their preferred outcome will be to remain in the trial till its completion. However, the participants in a trial may not only have expectations about the outcome. They may also care about the *experimental procedure* itself: the way the treatment are administered according to the trial protocol.

This point was first raised regarding randomization by James Heckman (1992) in a seminal paper about causal inference in field experiments. His argument focused on the non-comparative evaluation of social policy programs, where randomization simply decided who would join them (without allocating the rest to a control group). In this setting, we hypothesize that the outcome Y of participating in a program is causally determined by a given variable X, while participation in the program is itself determined by an umbrella variable Z. There is a set of values of Z, for which the subject will take part in the experiment. For instance, if we are evaluating a training program and, say, basic literacy is the outcome attained by the participants, we may presume it to be determined by their previous education, age, etc. Participation in the program may depend on certain values of income, employment, etc., all captured by Z. Heckman suggests we should distinguish between *regular participation in a program* and *participation in the program in an experimental regime*, where participation is randomized. Hence, we need to assume that:

a) Randomization does not influence participation

b) If it does, the effect is the same for all the potential participants.

c) Or, if different, it does not influence their decision to take part in the program.

Heckman's main objection is that randomization tends to eliminate risk-averse persons: if a potential participant wants to achieve the program's outcome but is only willing to take part in the test if she gets the treatment with probability 1, we will lose her for the experiment. This is only acceptable if risk aversion is an irrelevant trait for the outcome under investigation. However, even if irrelevant, it compels experimenters to deal with bigger pools of potential participants in order to meet the desired sample size, so the exclusion of risk-averse subjects does not disrupt recruitment. But bigger pools may affect in turn the quality of the experiment, if it implies higher costs. One way or another, argues Heckman, randomization is not neutral regarding the results of the experiment. Hence, the NIA does not hold.

In this view, the expectations of risk-averse persons about the *experimental procedure* (randomization) can generate a self-selection bias in the data: part of the potential population will not take part in the experiment if they become aware of randomization. According to Banerjee and Duflo (2009) and other advocates of field trials, we should blind participants about randomization in order to control their expectations about the procedure of the experiment. For instance, we can disguise randomization as a lottery by which the scarce resources of the program are allocated in order to make it more attractive to the participants. Or we may hide it straight away. As Banerjee and Duflo observe, "ethics committees typically grant an exemption from full disclosure until the end-line survey is completed, at least when the fact of being studied in the control group does not present any risk to the subject" (2009, p. 20). Participants in the experimental group will not know how they got involved and those in the control group may never know they have been excluded. If these latter are in different villages, as it often happens in trials run in developing countries, they may not get to know about the experimental treatment.

Summing up, the participants in a comparative trial may have expectations of two types, about the experimental outcome and about the experimental procedure. Expectations about the outcome may challenge the NIA and we should control for them with blinding devices. Expectations about the procedure may also bias the results (e.g., via self-selection effects, as we just saw) and blinding may also control for them, disguising or hiding the procedure. But, as we are going to see in the next section, there

are procedural expectations about being blinded that constrain our use of masking devices.

3. BLINDING AND DECEPTION IN ECONOMIC LABORATORIES

Let me know discuss procedural expectations in laboratory experiments in economics. Unlike in field experiments, in the laboratory the experimenter provides the environment in which the participants make their decisions. The laboratory is endowed with desks in which there usually is either pen and paper or computers that participants should use. They receive a set of instructions for the decision task. Depending on their performance, they will receive a monetary payment and the experimenter must ensure a direct and salient connection between decisions and payoff in order to make the results interpretable (or internally valid).

However, making such decision involves a *task construal* job for the participants: they do not have an algorithm to make them, just a set of instructions that they should interpret in the context of the experiment. And they will try to find additional information that may help them. Part of it may come from the social interactions in the laboratory and this may interfere with the goals of the experiment. The experimenter is in a position of authority regarding the participants, due to their procedural expectations: he knows what they should do. If the participants perceive anything the experimenter does as a cue prompting them to act in a certain manner, they may feel obliged to comply, generating an *experimenter's demand effect* (EDE). They would not behave as they would have done spontaneously, which is what the tested model aims at capturing. Let me quote one of the illustrations suggested by Zizzo (2010, p. 81):

> Chou et al. (2009) did a set of best guessing games experiments, with a standard rule that the winner would be the person guessing closest to 3/4 of the average guess by two players. One of their manipulations was to introduce a strong hint on how to play the game, by writing down (in bold characters, as a separate paragraph and with a figure on the side to stress the point) "Notice how simple this is: the lower number will always win". Chou et al. found that subjects largely followed this advice. In another manipulation they had what they label a "battle protocol" in which "your job is to choose how high to locate your troop on the hill, from 0 feet high to 100 feet high" and "you win the battle if your chosen location is higher than your opponent's". Subjects largely followed 'the job' they were given. Chou et al. interpret the findings as showing that the instructions enabled subjects to have a better game theoretical understanding of the guessing game.

There is another possible interpretation: there was a social EDE at work and subjects simply did what they were either strongly hinted or told to do. As such, Chou et al. may not be actually measuring the levels of reasoning of the subjects, which, given the abstract nature of the game, is arguably the main reason of interest of guessing games as such.

For the analysis of these interactions, Zizzo (2010) argues that we should consider the connection between the participants' expectations about the experiment's goals (and their subsequent actions) and the experimenter's true goals. He distinguishes three possible situations

(1)   Uncorrelated expected and true objectives

(2)   Negatively correlated expected and true objectives

(3)   Positively correlated expected and true objectives

In the first case, the true experimental objectives are obscure to the subjects, so their actions cannot engage in potentially confounding actions. In the remaining two cases, the participants' decisions can bias the experiment. An obvious solution would be to blind them regarding the true goals of the test, just as Duflo and her coauthors suggested about field experiments. After all, deceiving participants about the experiment they are in is an old tradition in fields such as social psychology. But deception has been traditionally proscribed in economic experiments.

In my view, the best rationale for this prohibition is grounded in an empirical analysis of the procedural expectation about blinding conducted by Ortmann and Hertwig (2002): *deception may generate biases*. After a systematic review of the literature on the various effects of deception in psychological experiments, they conclude that if participants suspect that they are being deceived, their behavior changes. In experiments on suspicion, the participants may receive tip-offs about the true purpose of the experiment, they may be made aware of experimental manipulation or they may be explicitly informed that they are being deceived. In all these cases, suspicion alters their performance in a wide range of variables experimentally investigated in social psychology. If they are simply told that some kind of deception might happen (or they are not told at all), there is no difference in their behavior compared to the control group. Moreover, direct experience of deception often affects the performance of participants in future experiments –though there is no evidence that indirect experience (e.g., word of mouth) has the same effect. For instance, participants

are concerned about being observed and judged, and they usually seek positive evaluations and avoid negative ones.

Blinding the true objectives of an experiment through deception can generate biases if the participants discover it. Zizzo (2010) recommends instead various forms of *non-deceptive obfuscation*, such as minimizing the interaction of experimenters and participants, using context-free language in the instructions or experimental designs that disguise the true goals of the experiment: e.g., arranging the experiment in various sessions so that participants forget about crucial details, incorporating "filler tasks", etc. In other words, we can only blind procedures partially. Full blinding would imply deception: lying to the participants about the true goals of the experiment from the informed consent form onwards.

The disclosure of the goals of the experiment creates a strategic interaction between the experimenter and the participants: in its simplest form, the experimenter can choose between being honest to the participants (about the settings and purpose of the experiments) or deceive them; the participants can choose between trusting the experimenter or question his claims (Ortmann and Hertwig 2001). It seems that we are sensitive enough to deception to the point of becoming suspicious once it happens, even if we do not have much to lose. This is a source of expectations about the experimental procedures that we need to control for, at the cost of partial disclosure. We need to trust the evidence showing that participants are less suspicious once they have been warned about the possibility of, at least, obfuscation and hope that this latter works.

In sum, we need blinding in order to enforce the experimental protocol and warrant the NIA. But the participants have procedural expectations about blinding: they do not like it if it goes as far as deception. Hence, blinding can only be partial and we therefore need to test whether it actually works. As we are going to see in the following section, drawing again on the experience accumulated in medical trials, the nature of these tests provides an additional argument for the necessity of partial blinding.

4. BLINDING AND DECEPTION IN CLINICAL TRIALS

Blinding was not uniformly used in clinical trials until the 1950s, but became a standard debiasing procedure when the FDA adopted the double blinded RCTs with placebo controls as their benchmark for testing the safety and efficacy of a compound. Placebos perform a double function in these trials: on the one hand, they conceal the

lack of treatment; on the other hand, they provide a baseline to determine the true efficacy of the experimental treatment compared to the natural rate of remission of the disease, including here the effects generated by the expectation of being treated. Of course, this baseline presupposes that blinding is effective and patients do not discover the actual treatment they receive. However, this has been an untested assumption until recently. As a matter of fact, the blinding procedures (and their success) are often imperfectly reported, despite the CONSORT recommendation in this respect. In view of my arguments in the previous section, I concur with this latter: we need to test the success of blinding in our experiments.

In order to test the degree of success of blinding, A. Hróbjartsson and coauthors (2007) investigated a sample of trials where blinding had been reported in comparison with another sample where there was no report –contacting the authors for further information. Let me quote their results:

> Thirty-one out of 1599 trials (2%) reported tests for the success of blinding. Test methods varied, and reporting was generally incomplete. Blinding was considered successful in 14 out of the 31 trials (45%) and unclear in 10 (32%). Of the seven trials (23%) reporting unsuccessful blinding the risk of a biased trial result was either not addressed or was discounted in six cases. We received 130 questionnaires from trial authors (65%) of which 15 (12%) informed that they had conducted, but not published, tests. (Hróbjartsson et al. 2007, p. 654)

The method used most frequently to test blinding was to ask patients to guess in which group they were. It was considered unsuccessful if their guess was better than random. Interestingly, according to Hróbjartsson et al. (2007, p. 657), in six out of the seven cases in which blinding was considered unsuccessful the authors either skipped or discounted the possibility of bias. The methodological quality of the trials where blinding was tested was similar to the trials in the sample.

There is no general agreement though about how to assess the success of blinding. Usually, participants are asked about their treatment guesses and their answers coded with different scales: e.g., they may choose between "placebo" and "active treatment" (and sometimes "don't know") or between a more nuanced qualification of their belief (such as "strongly believe the treatment is placebo", "somewhat believe…", etc.). There are two main approaches to the analysis of these answers. On the one hand, some focus on the correctness of the patients' guesses, usually through blinding indexes measuring their degree of disagreement with the treatment actually received. These indexes usually assign weights to each type of response. Correct guesses may point out

to unsuccessful blinding, but we need to interpret then the meaning of incorrect guesses and "don't know" answers. These latter, for instance, are considered the most supportive of success in the James' blinding index (Kolahi et al. 2009). However, we need to assume here that the patients are sincere and their responses do not constitute an experimenter's demand effect: they answer what they believe they are expected to. This is sometimes checked by requesting from the participants who give this answer a second "forced" guess between active treatment and placebo (Bang et al. 2004, p. 155).

As I mentioned before, unblinding may be caused by a true effect, so we may wonder at what point in the trial we should test for success. The indexes mentioned above usually rely on the participants' guesses at the end of the trial. But what if the proportion of correct guesses is the same at the beginning and the end of the trial: assuming that the allocation was properly concealed, shall we take it as a sign of unsuccessful blinding? The second approach to the problem focuses instead on the connection between the patterns of belief about treatment in each group throughout the trial and the clinical outcome under analysis. For instance, Rees et al. (2005) distinguish between *consistent*, *switching* and *undecided* beliefs, depending on how stable the guesses are at various points throughout the trials. In this second approach, the success of blinding is not measured by the correction of the guesses, but instead by their impact on the result: if patients with a similar pattern of belief (e.g., those who consistently believe themselves to be receiving a placebo) have a differential outcome compared to patients with other patterns, blinding will not be working properly.

In these two approaches, we find different conceptions of the placebo effect. In the former, having a correct guess seems necessary for our expectations to impinge on the trial outcome. In the latter, the placebo effect occurs when certain patterns of belief have a differential effect on the outcome. Whereas in the former, blinding aims at preventing the formation of correct guesses, in the latter blinding should block any connection between belief and outcome. If we knew the psychological/physiological mechanisms by which the expectations impinge on the treatment outcome we could try to block it directly. But we lack such knowledge, we may say we are blinding *blindly*, e.g., Rees et al (2005) found a lower rate of disease among those who consistently believed (correctly or incorrectly) that they were in the placebo group. Given that the active intervention was a water treatment unit installed at the kitchen sink, it seems

difficult to gather why believing that you do not have it installed led to lower rate of gastrointestinal infections.

As I said in the introduction, we have very little understanding of the causal effects of our expectations from treatments on the actual outcome, except that there seems to be some. Perhaps this is why the consequences of unsuccessful blinding are not often discussed. For instance, if the patients in the control group come to believe that they are receiving a placebo, the trial will compare a treated group with an untreated group in terms of the participants' expectations. Since we know little about the mechanisms by which these latter operate, the damage of unsuccessful blinding to the comparison is difficult to quantify[5].

| | | Participants told to expect: | |
		Active	Placebo
Participants get:	Active	Active treatment with expectancy (A)	Active treatment without expectancy (B)
	Placebo	Placebo with expectancy (C)	Placebo without expectancy (D)

Table 1 A balanced placebo design (Colagiuri 2010)

A potential solution would be to use an experimental design which factored out the net effects of expectation. But this would require deception, as the *balanced placebo design* shows. As illustrated in table 1 above, half of the patients in the active treatment will be told they will get a placebo (group B) and half of the patients getting a placebo will be told the opposite (group C). There is no blinding for the rest of participants: they know what they get, be it an active treatment or a placebo.

---

[5] As a matter of fact, it might be difficult to quantify even when blinding is successful if the placebo and the treatment responses are not additive: in this case, the placebo response in both arms may mask the true effect of the active treatment in one of them.

The virtue of this design is that we can factor out the effects of patients' expectation on the active treatment. But, of course, it assumes that patients will not suspect they are being deceived and exhibit a consistent pattern of belief about treatment throughout the trial. As we saw in the previous section, the more we use this design, the more we risk that suspicion spills over to future experiments. If patients suspect doctors to be lying to them, it will become increasingly complicated to test the success of blinding in this design through the patients' responses: e.g., a patient in group B might well think "I am told I have no treatment, but I feel better: has the doctor lied to me? If he did, does he expect me to tell him he's a liar?"

If patients take part in more than one trial, suspicion will inevitably spill over. In the experiments reviewed by Ortmann and Hertwig, indirect evidence of deception may not alter the participants' behavior. But, as I mentioned already, in these experiments there usually were not significant expectations about the outcome, as happens in clinical trials. With higher stakes in the outcome, patients may be more sensitive to deception: e.g., a meta-analysis has gathered evidence that when patients are randomized to their preferred treatment, the effect size is bigger than in patients who were indifferent to treatment assignment (Preference Collaborative Review Group 2008). Until we do not gather firm evidence on the effects of patients suspecting deception, we should apply the cautions against lying to patients in clinical trials that Ortmann and Hertwig recommended to experimental economics.

5. SHOULD WE ACCEPT THE NIA AT FACE VALUE IN FIELD TRIALS?

Let me summarize my argument so far: if the treatment outcome provides a good enough incentive for participants, they may violate the experimental protocol unblinding themselves, putting the NIA in question. The most efficient way to avoid it is to strengthen our blinding procedure, but only to a point where no suspicions of deception arise in the participants. If this latter happens, it may hinder any test for the success of blinding based on their responses. Hence, blinding is bound to be partial and it may well fail. However, this does not imply that we can dispense with blinding, since it is only warrant of the NIA: we need to test its success.

The experience accumulated so far in clinical trials shows that possible breaches in the NIA does not seem to diminish the external validity of their results (Teira & Reiss, forthcoming). However, the situation may be radically different in field trials,

because most interventions hinge directly on the expectations of the participants: they are performing tasks according to a set of incentives laid out by the experimenter. This is why the NIA requires a strong enforcement in a field trial for the conclusions to be credible. The participants' expectations should not be correlated in a systematic manner with the outcome of the experiment, just as in laboratory experiments, and we should not use deception if we are to test the success of our masking procedures. The development of effective blinding schemes is a serious challenge for the social trialist. But if they want a solid causal interpretation of their results, grounded on the NIA, there is no other alternative.

6. REFERENCES

Banerjee, Abhijit V., and Esther Duflo. 2009. The Experimental Approach to Development Economics. *Annual Review of Economics* 1 (1):151-178.

Bang, Heejung, Liyun Ni, and Clarence E. Davis. 2004. Assessment of blinding in clinical trials. *Controlled Clinical Trials* 25 (2):143-156.

Colagiuri, Ben. 2010. Participant expectancies in double-blind randomized placebo-controlled trials: potential limitations to trial validity. *Clinical Trials* 7 (3):246-255.

Colagiuri, Ben. 2010. Participant expectancies in double-blind randomized placebo-controlled trials: potential limitations to trial validity. *Clinical Trials* 7 (3):246-255.

Epstein, Steven. 1996. *Impure Science. Aids and the Politics of Knowledge*. Berkeley-Los Angeles: University of California Press.

Evans, Dylan. 2004. *Placebo : mind over matter in modern medicine*. New York, N.Y.: Oxford University Press.

Fergusson, Dean, Kathleen Cranley Glass, Duff Waring, and Stan Shapiro. 2004. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 328 (7437):432.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton.

Gluud, Lise Lotte. 15 March 2006. Bias in Clinical Intervention Research. *American Journal of Epidemiology* 163 (6):493-501.

Group, Preference Collaborative Review. 2008. Patients' preferences within randomised trials: systematic review and patient level meta-analysis. *BMJ* 337:a1864.

Heckman, James J. 1992. Randomization and Social Policy Evaluation. In *Evaluating welfare and training programs*, edited by C. F. Manski and I. Garfinkel: Cambridge and London: Harvard University Press.

Howick, Jeremy. 2011. *The Philosophy of Evidence-Based Medicine*. London: Wiley-Blackwell.

Hróbjartsson, A, E Forfang, MT Haahr, B Als-Nielsen, and S Brorson. 2007. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *International Journal of Epidemiology* 36 (3):654-663.

Kolahi, Jafar, Heejung Bang, and Jongbae Park. 2009. Towards a proposal for assessment of blinding success in clinical trials: up-to-date review. *Community Dent Oral Epidemiol.* 37 (6):477-484.

Ortmann, Andreas, and Ralph Hertwig. 2002. The Costs of Deception: Evidence from Psychology. In *Experimental Economics*: Springer Netherlands.

Petryna, Adriana. 2009. *When experiments travel : clinical trials and the global search for human subjects*. Princeton: Princeton University Press.

Rees, Judy R., Timothy J. Wade, Deborah A. Levy, Jr. Colford, John M., and Joan F. Hilton. 2005. Changes in beliefs identify unblinding in randomized controlled trials: a method to meet CONSORT guidelines. *Contemporary Clinical Trials* 26 (1):25-37.

Schulz, Kenneth F., Iain Chalmers, and Douglas G. Altman. 2002 The Landscape and Lexicon of Blinding in Randomized Trials *Annals of Internal Medicine* 136 (3): 254-259

Schulz, Kenneth F., Iain Chalmers, Richard J. Hayes, and Douglas G. Altman. 1995. Empirical Evidence of Bias. *JAMA: The Journal of the American Medical Association* 273 (5):408-412.

Schulz, Kenneth F, David A Grimes, Douglas G Altman, and Richard J Hayes. 1996 Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology *BMJ* 312 (7033):742-744

Teira, D., and J. Reiss. forthcoming. Causality, Impartiality and Evidence-Based Policy. In *Mechanism and causality and biology and economics*, edited by H.-K. Chao, S.-T. Chen and R. Millstein. Berlin: Springer.

Zizzo, Daniel. 2010. Experimenter demand effects in economic experiments. *Experimental Economics* 13 (1):75-98.