

Limitations of the Double-Blind Pharmaceutical Study:

*Vernon M Neppe MD, PhD, FRSSAf, DFAPA, BN&NP, MMed, DPsM, FRCPC, FFPsych^a
Director, Pacific Neuropsychiatric Institute, Seattle, WA, USA
(Adj. Full) Professor, Dept of Psychiatry, St Louis U., St. Louis, Mo*

Abstract:

This paper examines the limitations of the double-blind pharmaceutical study in medical research. These areas are often neglected. First, common difficulties are discussed. These include lack of appropriate demographic controls including cigarettes, alcohol, caffeine and drug interactions. Additionally, not eliciting side-effects and therapeutic effects may occur because the measuring instruments are insufficient. Additionally, insufficient duration of the study may lead to inappropriate interpretations of efficacy or lack of efficacy, insufficient subject size may produce insufficient power, recruitment difficulties may produce distorted populations, crossover studies may result in lingering effects and withdrawal confounding factors, and statistical aberrations may lead to the wrong conclusions but particularly certain drugs are not necessarily suitable for such methodologies because of their special variable dose requirements. All these may lead to Type 1 and 2 errors in analysis. The subtle difficulties of blind studies are then examined: The experimenter effect; the triple-blind study; intelligent prescription; the distortions of published versus non-published studies; and adjunctive medication. These factors all exemplify the challenge of appropriate methodology that can be more adequately interpreted.

Keywords: *adjunctive medication , alcohol , analysis , appropriate methodology, caffeine , challenge , cigarettes , confounding factors, crossover studies , demographic controls , distorted populations , distortions, double-blind , drug interactions , duration of the study , efficacy , experimenter effect , inappropriate interpretations , incorrect conclusions , intelligent prescription , interpretation , limitations , lingering effects , measuring instruments , medical research, methodologies , neglected areas , pharmaceutical study , power , published versus non-published studies , recruitment , side-effects , specific drugs , statistical aberrations , statistics , subject size , subtle difficulties , therapeutic effects , triple-blind study , Type 1 errors, Type 2 errors , variable dose , withdrawal.*

Essential background

^a Dr Neppe is a psychopharmacologist, behavioral neurologist, neuropsychiatrist, forensic specialist, psychiatrist and epileptologist. He is author of seven books, including the classic “Cry the Beloved Mind: A Voyage of Hope” (see www.brainvoyage.com).

A blind study is a clinical trial in which the subject or the investigator (or both) are unaware of which trial product/drug the subject is taking. When only one of them is blind to that data this is a single-blind study and when both don't know which treatment a subject is receiving, the study is double-blind.

The double-blind, controlled medication study (DBCMS) has become a standard in medical research. Essentially, the study is double-blind because the patient doesn't know what he or she is receiving, nor does the physician or ranker doing the ratings. In this way, both patient and ranker are blinded and therefore misconceptions or prejudices are supposedly eliminated.

double-blind studies are the standards by which drugs are approved. Indeed, within the United States, the FDA [Federal Drug Administration] generally requires two double-blind studies showing the drug is superior to placebo and at least equal to a standard other competing drug indicated for the particular condition in which the drug is being studied.

There are two major kinds of DBCMS. The more common one by far is the "between patient study" (BP) (where patients are randomized into two groups). In the rare "within patient crossover study" (CO) the patient randomly receives initially either placebo or active drug and then is "crossed over" to the alternative they did not receive— either active drug or placebo. (e.g. Neppe ¹)

Many DB studies do not involve placebo but one standard already approved medication compared with the new drug to be studied. DB studies may also involve all three arms —placebo, active drug, and standard approved comparative drug for that condition. This three arm study ², or for that matter, blindness to dosage, remains by definition double-blind (DB), despite some mistakenly calling them "triple-blind" e.g. Jensen ³). This is because neither the patient nor rater knows the drugs being used but the experimental protocol leader still does. By contrast, a triple-blind study would imply that none of the patient, rater or person uncovering the code in the analysis can identify who is taking what. Triple-blind studies are seldom performed because of the complexity and the fact that they are not regarded as necessary in medicine.

But is it? This and other confounding factors are discussed below. For the uninitiated in this area, the comments below may appear obvious. And yet they are repetitively ignored and therefore may cast a shadow upon our studies of medically approved medications.

Common difficulties with the double-blind study

Lack of adequate demographic controls

When studies are BP (between patient), the patients are randomized such that essential demographics, such as age, sex, relevant facets of health such as blood pressure or weight, and sometimes racial/ethnic group are controlled for.

However, in most studies, remarkably, other essentials, such as use of alcohol, cigarettes, and caffeine, as well as nutritional supplements are ignored unless the study specifically involves smokers or alcohol users as key variables or confounding factors. Yet these may have major impacts on the results of other studies, particularly those on subtle brain functions such as depression. Remarkably, in a PubMed search (December 2007), I could not find a single study that definitively ensured that patients were controlled for these three basic variables, or were excluded entry into the study when using significant alcohol or caffeine or using cigarettes. This is amazing because alcohol, cigarettes and caffeine, *inter alia*, may potentially bias studies confounding personality structure, addiction potential, interactions with medications at the metabolism level (e.g. liver —pharmacokinetics) and at the receptor interaction level (e.g. brain pharmacodynamics).

Cigarettes, for example, may subtly speed breakdown of certain chemicals by some of the polycyclic aromatic hydrocarbons interacting at the 1A2 level of the P450 cytochrome enzyme system level in the liver ⁴. They may also cause major interactions due to the nicotine related effects. ⁴. Similarly withdrawal may play a role. ⁵. And this does not take into account controlling for the vast number of interactions and potential distortions we encounter because the majority of outpatients in Medicine today (in the United States) are using nutritional supplements. Nutritional, herbal and alternative medications, remarkably, have not been well studied—and would possibly encounter significant under-reporting to physicians of this phenomenon as patients often feel reluctant to share such data with their physicians, or don't realize the importance of doing so.

Non-elicitation of effects or side-effects

Unfortunately, for each added variable that is controlled for, the power of the study diminishes because equivalent samples diminish in size. The hope in many studies is that potential uncontrolled confounding factors would “wash” out after randomization. But without directly at least eliciting such data (even if it is not specifically controlled for directly), we cannot demonstrate that these variables are, in fact, not significant. Even more so with pharmaceutical sponsored studies, the researchers are often paid to evaluate an already defined multicenter, pharmaceutical company authored, specific protocol. Whereas these protocols invariably allow for eliciting added side-effects, many patients will not report them spontaneously, and even more so the studies usually have non-physician coordinators who may not be astute enough to detect such changes. So the extra symptoms are not detected. Moreover, the physicians in charge of the coordinators may, at times, see the patient only briefly. The classic example in this regard relates to how several of the Selective Serotonin Reuptake Inhibitor antidepressant drugs were not initially noticed as causing profound libidinal loss because patients in drug studies don't spontaneously report sexual problems. We now know that diminished libido is so common with these medications that it is the exception for the patient not to have this side-effect.

Insufficient duration of the study.

Very often, these studies— DBCMS—are too short term. A classic example is from psychiatry: Most antidepressant studies are only six to eight weeks long, looking at acute management of depression. The assumption is that these drugs would maintain their effects. This short duration was based on the main original group of antidepressants, the tricyclics that empirically seemed to maintain their effects for decades so it was felt that there was no need to look long-term at the maintained efficacy of such drugs. This assumption has been demonstrated to be false. Indeed, there is only one recently approved drug, venlafaxine hydrochloride, which has been demonstrated to maintain patients at remission prophylactically for periods of two years.⁶ Studies particularly with a major group of antidepressant drugs, the most popular of the 1990s, the Selective Serotonin Reuptake Inhibitors (SSRIs), have demonstrated efficacy in maintaining remission that is better than placebo at one year e.g. fluoxetine⁷, paroxetine⁸, sertraline⁹, but when studied, have not shown maintained efficacy at two years (e.g. paroxetine and sertraline)?^{6, 7} As an aside, large two-year multisite studies are not routinely performed, indeed they are rare as they cost a fortune so this does not necessarily mean to say that these drugs do not work for this period of time; it's just that this has not been demonstrated—however, there is some cogent reasoning why there could be loss of efficacy over time for these SSRIs because these drugs work on serotonin and not on norepinephrine, and they may cause too much reuptake inhibition producing compensatory mechanisms from the brain⁴

The crossover study

Most studies are BP because CO studies are difficult to perform. When they are purely placebo controlled, within patient CO studies, the same patient receives both the active drug and placebo in different phases. This assists with the demographic contradictions that may occur in the between patient (BP) studies, but instead the protocol needs to sort out another problem — if the patient has received the active drug first, is there a continuation effect of that drug continuing into the next placebo phase, or, alternatively, is there a withdrawal effect? This cannot be easily solved by always maintaining the placebo phase first, as then this would no longer be double-blind.

Recruitment and size

Another problem of the DBMCS is adequate recruitment of subjects. Imagine being part of an investigation of complementary cancer treatment. Would you participate when knowing that your treatment would be determined by randomization and you may just die if you're in the wrong group? Rostock and Huber had to abandon their study given that patients preferred to be able to actively choose which drug they wanted to receive.¹⁰ Their study was abandoned because they could not recruit a large enough sample.

Statistical aberrations.

Statistically, studies are analyzed using parametric or non-parametric statistics with a predefined level of confidence (generally, one in twenty against chance). To demonstrate superiority, the sample size must be adequate, particularly when studies involve very similar compounds. This creates easy potentials towards results that may not be appropriate. Similar compounds may be regarded as equal in efficacy because the statistical power is insufficient to differentiate them. Alternatively, the sample size may be insufficient to demonstrate real differences.

These Type 1 and 2 errors are the tip of the iceberg. Imagine being able to demonstrate that in a very large sample, drug A (being tested) is statistically better than drug B. It may be that drug A helped only 75% of patients and drug B helped 73%. And this difference may be in practice inconsequential: These statistics may not correlate with real clinical differences. We in reality want to be able to demonstrate vast clinical superiority of one drug over another, not just clinically insignificant though statistical superiority.

Subtle difficulties with the double-blind study

There are even more difficult and insidious problems to controlled double-blind studies.

Predictability of the drug's identity.

The challenge here is the following: Are double-blind studies really double-blind?

There certainly may be logical difficulty differentiating the active drug being studied versus a similar already indicated standard compound because their clinical benefits and side-effects may be similar. However, in some studies, particularly placebo controlled ones, the labeling of a study as double-blind would require a level of clinical acumen in the physician equivalent to a layperson, as the side effects or extent of response would be dramatically different. For example, some of the group of drugs called beta blockers, such as propranolol and nadolol, actually slow the pulse significantly as part of their therapeutic action. You cannot call such studies double-blind, when immediately a good physician examining the patient would feel the pulse and be able to know that highly probably the drug that is being used is the active drug, and if the pulse were not slowed and yet the patient was taking sufficient dosage, the drug they are likely is placebo. This leads to the researcher potentially being biased by not being blinded.

At a subtler level, predictions can be made by patients with some accuracy; For example, a recent study on mistletoe suggested that, even with nutritional substances, in 100% of cases the drug could be predicted when the patient took the active drug and in 70% of cases could be predicted when placebo was taken. But this was easy as it was based on a skin reaction side-effect. ¹⁰

The experimenter effect.

The experimenter effect has been largely ignored in medical research ¹¹, but it has been repeatedly demonstrated to be of relevance and well described in the parapsychological domain, extending gradually into medicine and into psychology. Effectively, what happens is the experimenter influences the results of the study. They influence the results because they impact on the data occurring modifying outcome because subtle other effects occur. The most extreme variation may be "doctrinal compliance" ¹², where the experimenter so seems to influence the study that the results comply with his expectations.

Alternatively, researchers may particularly attempt to eliminate such psi biases e.g., Green,¹³ Sometimes this need not have any invocation of alleged psi. In its simplest form, imagine a physician who was very kind, caring, considerate, insightful, and providing a great therapeutic environment for the patient. We could imagine patients doing very well under those circumstances even when taking placebo: Therefore, despite the randomization, there may be no statistical correlation because the researcher was too good and influenced the outcome positively. Similarly unconscious biases—beliefs and expectations— may lead to different behaviors by the researchers.¹⁴ This is a major issue.^{15, 16}

This kind of flaw in research has led to the triple-blind study. There are unfortunately very few real Triple-Blind Studies. One example is a novel psychological protocol reported this year where effectively there were blind experimenters, blind data collection, blind subjects as to the real data being researched.¹⁷ Realistically, these are very difficult to implement.

Intelligent prescription.

Another confounding issue is that some drugs do not do well in the double-blind paradigm because this DB paradigm implies that the best or correct dose for the condition is obvious and easily known and consistent across patients. However, this is not always so: Take, for example, the drug buspirone. I sometimes have referred to this medication as the “intelligent psychiatrist’s medication.” This is one of the most remarkable compounds ever developed¹⁸, and yet, many will say, “This is just placebo. It doesn’t work.” In fact, in my experience, based on tens of thousands of contacts of physicians and their patients around the country, the drug works extremely well when used appropriately. If it is appropriately prescribed in the correct doses, it can work remarkably, for example as an adjunct to methylphenidate in attention deficit disorder¹⁹

Yet, this drug, at times, has failed to show an effect in double-blind studies. Effectively, double-blind studies compromise its most effective use. This may be because the dose of the drug is critical because it has neurotransmitter effects at different doses and the dose, at times, requires good clinical acumen to pinpoint. This is one reason, for example, why one of my studies (tardive dyskinesia) was done in a single-blind context.²⁰ The patient knew they were on the active drug. They were uncertain of the dosage and were not aware of more details. However, the raters were unaware of the dosage being used, which could have been anywhere from zero to very high. They were blind as to dosage and so were the patients but as with almost all the clinical studies, the blind was broken by the experimenters.²⁰ Such single-blind studies may allow appreciation of the virtues of a drug that may be missed.

Published studies

I mentioned how some drugs may fail in DBCMS. Such studies are sometimes not even published or abandoned before completion and the literature becomes an empty hole in that area. Therefore when meta-analyses are then done of published studies, the results may be positive when they should have generated chance probability. Often the pharmaceutical company sponsoring the study has the final say on its publication and do not allow it and the researchers are contractually obligated to not discuss it. In similar vein, if sponsors control studies, they may just submit the positive studies to a regulatory body such as the FDA and therefore the data submitted may reflect statistical significance, whereas if all the studies were combined the diluted effect of negative studies would produce non-significant results.

Moreover, such sponsors may not want to study a particular feature: Do you think a company that has spent several hundred million dollars marketing a hypnotic to assist sleep wants to do a further study demonstrating that their drug may cause on subtle testing impairments in responsiveness an hour after waking? Some studies do not get done because they are most commonly not independently sponsored—and researchers don't have millions to do such studies without such funding.

Adjunctive medication

Finally, there are significant ethical issues to drug studies. I personally find it very difficult to study a drug versus placebo when I know that that drug may actively work far better than placebo. How can I justify giving my patients placebo? This may lead to an area not being studied. An example is my research with tardive dyskinesia, an incurable condition that could be "cured" possibly by buspirone in very high doses.²¹ We could not perform a double-blind placebo controlled study because I deemed it unethical to do so and indeed our experience over the past nearly two decades appears to justify that impression.

This brings about a whole new ballgame of using adjunctive medication. This appears a legitimate but underused technique: adding the drug to the particular patient's medication regime, with ratings beforehand and afterwards. This may well be better served by within patient crossover studies¹ as opposed to these between patient studies. The applicability of adjunctive medication should be taken into account because this reflects the greater component of the real patient and the real patient's condition.¹

The challenges

These elements suggest that there are limitations to our current ways of evaluating drugs. Well-controlled DB studies do indeed have ostensible patient and rater blindness but the role of the experimenter effects and the abilities to differentiate active drug particularly from placebo confound this. Moreover, the durations of the study may be insufficient. Thirdly, there may be problems with randomization because studies are not currently controlled for alcohol, cigarettes, caffeine, and nutritional supplement usage, as well as other possibly relevant confounding factors and the elicitation of side-effects may not be optimally obtained by the structured protocol.

Next, there are problems with drugs that require careful dosing which blinded studies cannot easily achieve. Furthermore, designs such as choosing cross-over versus randomized between patients need be carefully considered. Additionally, the limited number of adjunctive controlled studies may be unfortunate, as are the ignored negative studies. Sample size and the differentiation of real clinical effects are other challenges.

Now, surely, many of these limitations are obvious? Yet ironically, we see these problems—even the more easily reparable ones—repetitively today in our double-blind studies.^{22, 23, b}

1. Neppe, VM. Carbamazepine as adjunctive treatment in no epileptic chronic inpatients with EEG temporal lobe abnormalities. *J Clin Psychiatry*. 1983, 44:9, 326-331.
2. Heller, A, Zahourek, R, Whittington, HG. Effectiveness of antidepressant drugs: a triple-blind study comparing imipramine, desipramine, and placebo. *Am J Psychiatry*. 1971, 127:8, 1092-1095.
3. Jensen, R, Brinck, T, Olesen, J. Sodium valproate has a prophylactic effect in migraine without aura: a triple-blind, placebo-controlled crossover study. *Neurology*. 1994, 44:4, 647-651.
4. Neppe, VM. *Cry the Beloved Mind: A Voyage of Hope*. Seattle: Brainquest Press (with Peanut Butter Publ. Publishing). 1999.
5. Marzilli, TS, Hutcherson, AB. Nicotine deprivation effects on the dissociated components of simple reaction time. *Percept Mot Skills*. 2002, 94:3 Pt 1, 985-995.
6. Keller, MB, Trivedi, MH, Thase, ME, Shelton, RC, Kornstein, SG, et al. The Prevention of Recurrent Episodes of Depression with Venlafaxine for Two Years (PREVENT) Study: Outcomes from the Acute and Continuation Phases. *Biol Psychiatry*. 2007, 62:12, 1371-1379.
7. Kocsis, JH, Thase, ME, Trivedi, MH, Shelton, RC, Kornstein, SG, et al. Prevention of recurrent episodes of depression with venlafaxine ER in a 1-year maintenance phase from the PREVENT Study. *J Clin Psychiatry*. 2007, 68:7, 1014-1023.
8. Montgomery, SA, Dunbar, G. Paroxetine is better than placebo in relapse prevention and the prophylaxis of recurrent depression. *Int Clin Psychopharmacol*. 1993, 8:3, 189-195.

^b This article was originally written for Telicom, the Journal of the International Society for Philosophical Enquiry, www.thethousand.com. It has been slightly adapted including additional abstract, keywords and references. See reference 23. © Vernon M Neppe.

9. Lustman, PJ, Clouse, RE, Nix, BD, Freedland, KE, Rubin, EH, et al. Sertraline for prevention of depression recurrence in diabetes mellitus: a randomized, double-blind, placebo-controlled trial. *Arch Gen Psychiatry*. 2006, 63:5, 521-529.
10. Rostock, M, Huber, R. Randomized and double-blind studies--demands and reality as demonstrated by two examples of mistletoe research. *Forsch Komplementarmed Klass Naturheilkd*. 2004, 11 Suppl 1, 18-22.
11. Neppe, VM. The experimenter effect in medical research. *South African Medical J*. 1982, 62:3, 81.
12. Ehrenwald, J. Cerebral localization and the psi syndrome. *J Nerv Ment Dis*. 1975, 161:6, 393-398.
13. Green, PR, Thorpe, PH. Tests for PK effects in imprinted chicks. *Journal of the Society for Psychical Research*. 1993, 59:830, 48-60.
14. Forster, KI. The potential for experimenter bias effects in word recognition experiments. *Mem Cognit*. 2000, 28:7, 1109-1115.
15. Kirk-Smith, MD, Stretch, DD. Evidence-based medicine and randomized double-blind clinical trials: a study of flawed implementation. *J Eval Clin Pract*. 2001, 7:2, 119-123.
16. Behi, R, Nolan, M. Experimental designs. *Br J Nurs*. 1996, 5:12, 754-756.
17. Beischel, J, Schwartz, GE. Anomalous information reception by research mediums demonstrated using a novel triple-blind protocol. *Explore (NY)*. 2007, 3:1, 23-27.
18. Neppe, VM. *Innovative Psychopharmacotherapy*. New York: Raven Press. 1990.
19. Neppe, VM, Young, Z. Buspirone as a new treatment for attention deficit disorder and aggression in children and adolescents. *Australian J of Psychopharmacology*. 2003, 10: September, 47-56.
20. Moss, LE, Neppe, VM, Drevets, WC. Buspirone in the treatment of tardive dyskinesia. *J Clin Psychopharmacol*. 1993, 13:3, 204-209.
21. Neppe, VM. High-dose buspirone in case of tardive dyskinesia. *Lancet*. 1989, 2:8677, 1458.
22. Neppe, VM. Double-blind Studies in Medicine: Perfection or Imperfection? *Telicom*. 2007; 20: 6, Nov. -Dec., 13-23
23. Neppe VM. Ethics and informed consent for double-blind studies on the acute psychotic. *Medical Psychiatric Correspondence: A Peer Reviewed Journal, Model Copy*, 1990; 1 (1): 44-45.