Are we blind to the limits of double-blind medical studies?

Opinion

Volume 5 Issue 6 - 2016

Vernon M Neppe*

Department of Neurology and Psychiatry, St. Louis University, USA

***Corresponding author:** Vernon M Neppe, Director, Pacific Neuropsychiatric Institute, Seattle, WA and Executive Director, Exceptional Creative Achievement Organization and Adj. Professor, Department of Neurology and Psychiatry, St. Louis University, USA, Email: psyche@pni.org

Received: May 03, 2015 | **Published:** May 10, 2016

Opinion

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 [1,2]. In Medicine, scientists and clinicians regard one specific kind of blind study, the double-blind (DB) study-in which both patient and rater are unaware of the medication the patient is receiving— as *the* definitive way to prove a drug is useful. Let us examine some limitations of DB studies: Based on the features below, the reality is that DB studies may sometimes be flawed, or may be largely irrelevant clinically. Moreover, other studies that are not DB may be clinically far more useful.

There are different kinds of blind studies (Table 1)

Table 1: Kinds of blind study.

a)	**Unblinded study:** A study in which both the patient/ subject and the doctor /investigator knows what is being administered. It is far the most common study in psychology, for example, where tests and scoring is obvious to the subject and the tester. It is also the way physicians practice medicine in a clinical setting, except they may not perceive such practice as "research", but a clinical attempt to make the patient better. There is a special psychology involved here and the doctor-patient relationship may play a key role. But the object is to get the patient better, though we still might not know how well the drug works compared with the psychology of the relationship between the therapist or doctor and the patient.
b)	**Single blind studies:** When only one of the subject of patient OR the investigator is blind to the data being examined, this is a single-blind study. Sometimes the rater knows what the patient is taking (e.g. placebo or active drug) but is blind as to other data (e.g. the dose). This is still single blind: Such dose adjustments are important variants because the research can demonstrate that there may be special optimal doses for specific interventions (e.g. not too high or too low; or that side-effects to therapeutic effects preclude particular dosing).
c)	**Double-blind:** In DB studies, neither the patient knows what he or she is receiving, nor does the physician or ranker doing the ratings know. In this way, both patient and ranker are blinded and therefore misconceptions or prejudices are supposedly eliminated. This is by far the most common research study done in clinical medicine, because it achieves (with a properly performed study) a rather definite indication that the intervention (e.g. the medication) works more than by chance. But it does not indicate that the patient may benefit markedly from the drug: The result may be only marginal not clinically significant effects.
d)	**Triple-blind studies:** In triple-blind studies. none of the patient, rater or persons uncovering the code in the analysis can identify who is taking what. Triple-blind studies are seldom performed because of their complexity, and the fact that they are not regarded as necessary in medicine. Sometimes researchers regard this term "triple-blind" incorrectly: e.g. blindness to dosage, remains by definition double-blind (DB), despite some mistakenly calling them "triple-blind"—this is because neither the patient nor rater knows the drugs being used, but the experimental protocol leader still does making it DB.
e)	**Quadruple-blind studies:** the subjects, investigator(s), evaluator(s), and the data analysts all remain blinded. This is very difficult to perform and complex, so it is rare, but could be used sometimes in Consciousness Research where no-one at the time of the study knows the answer which may remain, for example, on a computer where the data is being / will be generated.

The *double-blind,* controlled medication study (DBCMS) has become a standard in medical research. In the United States, the FDA (Federal Drug Administration), approval of a new treatment often generally requires two double-blind studies showing the drug is superior to placebo, and at least equal to a standard other competing drug that is indicated for the particular condition in which the drug is being studied. This involves statistical differences (such as $p<0.01$ or at times $p<0.001$). It also requires appropriate safety data.

There are two major kinds of placebo related DBCMS.

(i) By custom, possibly, by far, the more common research study, is the "between patient study". In this, the patients are randomized into two groups, active drug or placebo. This may be a problem, e.g. with dangerous conditions: Would you like to have an incurable cancer and be assigned blind to a "placebo" group, as opposed to having a "chance" with an active new intervention drug?.

(ii) 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 [3-5]. In between the crossover period, there may or may not be a "washout" period with placebo. CO studies are often neglected though useful, because it allows far fewer subjects, and ultimately the patients have some knowledge of whether the drug will help them specifically. Moreover, it might lead to a continuation study where the patients who are responding can benefit from the drug.

There are sometimes obvious problems with purely placebo controlled DBCMS research listed in Table 2 and amplified below.

Table 2: Some limitations of double-blind studies.

1. Not blinded to the rater.
2. Not blinded to the patient.
3. The "wrinkled paper fallacy" of many studies is done, only positive ones being reported.
4. The methodology
5. Sampling the population problems
6. Sampling the data problems
7. Raters of the data problems
8. Non-elicitation of side-effects or therapeutic effects.
9. Confounding elements problems
10. Ethics of the study
11. Politics

a) Actually not blinded to the rater: Sometimes the study, though purporting to be DB, is not effectively "blinded" to the rater. This because based on clinical response or side-effects, the astute clinician rater can, with relatively high probability, predict whether the patient is on the active intervention, not the placebo. This is sometimes easy as the efficacy shows obvious changes (e.g. beta-blockers slow the pulse) or side-effects may give great clues. This might allow unblinding with great accuracy [1,2]. This is not always so. Many DB studies do *not* involve placebo but, instead, use one standard already approved medication compared with the new drug to be studied. DB studies may also involve all three arms-the placebo, the active drug, and a identical-looking standard drug that has already been approved comparative for that condition. This would, then, be a three arm study [2]. In that instance, it may be more difficult to predict, for example, whether the patient is taking the experimental drug or the standard.

b) Not blinded to the patient: Sometimes the study though purporting to be double-blind, is not blinded to the patient: In that kind of instance, the patient can also postulate with great accuracy whether they are receiving active drug or placebo: The patient, for example, might have side-effects, and can guess, based on their previous and current experience, that they're on active drug. Alternatively, the patient may improve so much on the "new antidepressant" that they're reasonably certain what arm of the study they're on. They might be wrong, and this may be purely a "placebo" response, but certain patients are usually astute enough to make correct interpretations. This biases the research. Now in both instances, the interpretation might not be certain but it partly unblinds the research because the object of DB studies is not to have opinions that might prejudice.

c) Wrinkled paper fallacy: Moreover, the results of the study might not correctly show true statistical effects: The study, for example, the overall interpretation of the studies might be flawed because not all results have been released: The company sponsoring the research, understandably, may want to show positive results. Consequently, the pharmaceutical company or sponsoring research group may have performed other studies but submitted only those that were positive. They sometimes might rationalize that there were flawed errors in the data of the rejected data-and, indeed there might have been. This creates the "wrinkled-paper fallacy" where analyses of all studies may not have produced the same results as those that were submitted: For example, only two results reflecting a 1 in 100 against chance result would be different from six studies, where four were discarded as not significant and not pertinent: When all six studies would be pooled, they might be overall not significant statistically. In some jurisdictions, there are attempts to demand all data be released. Such release would be excellent to maintain objectivity, and if needed the flaws of each study can be pointed out.

d) Inappropriate methodology: It is remarkable how often major studies are planned by MDs who do not have PhDs. I have repetitively seen expensive studies (hundreds of millions of dollars worth) being ruined because proper research methodologists with clinical insights were not employed in planning the actual bones of each step of studies. There is an enormous difference between MDs (almost all of whom have not received formal training in methodology of research) and PhDs (who have the research methodology and should be involved in every study). Optimally, the MD, PhD with proper training in that discipline should be used. This kind of study may preclude positive results because the criteria were incorrect, or may over diagnose conditions. This might mean the criteria for patient selection for the study might have been compromised.

e) Sampling the population problems: Studies generally require specific admission criteria. Sometimes some facilities will admit say 90% of applicants with a specific diagnosis, while other facilities may regard only 5% as appropriate for the study. This kind of conflict happens in my experience, although the admission criteria are the same. Such stringency differences might lead to different outcomes as some patients should not have been admitted based on the diagnoses, or alternatively, they may be excluded unjustifiably. This also distorts statistical analyses.

f) Problems sampling the data: Criteria at each level need to be defined carefully to prevent error. This can easily be misinterpreted particularly as data in medical research is almost always "ordinal" meaning lists such mild, moderate, severe are sometimes subjective.

g) Raters of the data: Often the raters are not adequately trained for ranking symptoms. Some studies stipulate inter-rater reliability criteria. However, evaluating many patients in an hour might limit the success of such rankings. There appear to be times when facilities employ Bachelor's-level individuals to rank, and the MD signs off after seeing the patients only cursorily for a far shorter period than he/ she should have. This might compromise rankings.

h) Non-elicitation of side-effects or therapeutic effects: Studies are only as good as their protocols. There are many examples, some mentioned briefly above.

i. Absence of clinical effects may occur because the measuring instruments are insufficient. Sometimes sensitivity is an issue. For example, the AIMS is often used in tardive dyskinesia, where the STRAW is far more sensitive [6]. But, on the other hand, the STRAW is unproven and not standard [7].

ii. Additionally, insufficient duration of the study may lead to inappropriate interpretations of efficacy or lack of efficacy. For example, double-blind studies of six or eight weeks may demonstrate efficacy, but it doesn't mean that there may be maintained effects over years. The loss of efficacy we see with the SSRI drugs is an illustration [8].

iii. Insufficient subject size may produce insufficient power.

iv. Recruitment difficulties may produce distorted populations, as indicated.

v. Crossover studies may result in lingering effects or withdrawal, confounding factors. This may be the reason for washout periods in between, but still the effects may linger or withdrawal may be pertinent.

vi. Statistical aberrations may lead to the wrong conclusions. And:

vii. Critically important, is that certain drugs are not necessarily suitable for such DB methodologies, because of their special variable dose requirements and specialized individualized prescriptions. Buspirone is an obvious example [9].

All these points, might lead to Type 1 and 2 errors in analysis [1].

i) Non-elicitation of effects or side-effects: Patients sometimes consciously do not want to report side-effects. This way they feel they will not be "dropped from the study". Also symptoms, such as sexual problems or incontinence of urine, may be embarrassing to mention. Ignoring symptoms by the patients are one side; and not writing in certain symptoms like sexual problems into the protocols also can lead to dramatic underreporting, such as with selective serotonin reuptake inhibitors [8]. The classic example in this regard relates to how several of the Selective Serotonin Reuptake Inhibitor antidepressant drugs were not initially apparently 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 [8].

j) Confounding elements: There are subtle, often ignored difficulties of blind studies: These include:

(a) The experimenter effect-the influence of those involved in the research [10-13].

(b) Intelligent prescription-the clinician's awareness of dosage to prescribe-is much more difficult in DB research. Moreover, the patient might have responded if the correct dose could have been chosen.

(c) The distortions of published versus non-published studies, is important: This may because the studies have been rejected by journals after they've been submitted. This may not have been published because the studies did not yield significant statistical results.

(d) Adjunctive medication and supplements are sometimes ignored but may be key factors. Often, for example, the roles of cigarettes, alcohol, nutritional supplements, pain prescriptions, or other medications taken as needed, and recreational drugs (often not admitted to) play important roles.

(e) Other environmental events such as stress, exercise, travel, moving across time zones, and poor medication compliance are often unmeasured factors.

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 they are not specifically controlled for directly), we cannot demonstrate that these variables are, in fact, not significant. These further factors exemplify the challenge of adequately interpreting data based on appropriate methodology. Unfortunately, for each added variable that is not controlled for, the "power" (the potential ability of the research to generate statistically significant results) of the study diminishes.

a) Ethics: A study where the comparative drug is the best available medication approved for that diagnosisså as comparison has a major ethical advantage, because patients are receiving the experimental agent or a known good treatment for that condition.

b) Politics: 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, sometimes just under "other", 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 most important point

So far, I've discussed the problems within double-blind studies. The major motivation for this editorial is this final point: Sometimes non-blind or single blind studies are better. A classical example is the Neppe work on high-does buspirone in tardive dyskinesia [6,7,14]. In this example, the dose is critical and knowing what one is doing is very important for the success of treatment.

But the other aspect is it does not matter if the best double-blind study shows statistical significance if it is not clinically relevant (Table 3). We clinicians do not want our patients on antibiotics to respond only in 50% of cases even if that is pertinent in research studies because it might be better than placebo. If we have found the bacteria involved, we should expect at least a 95% clinical result to successfully treat an uncomplicated patient with a bacterial infection. However, sometimes, as in cancers, our gauge would be different: a 50% success rate may be great!

Table 3: Clinical versus research significance in Medicine are quite different.

(i)	**Research significance:** Statistical studies showing significance-demonstrating that the effect of the investigational drug is better than the controlled effect of placebo, and / or, at least, equal to the effect of the best available treatment approved for such a condition: Research significance demonstrates that the "null hypothesis" of no relationship between variables has been actively refuted.
(ii)	**Clinical significance:** Clinical significance refers to the practical relevance in the medical and psychological areas of a treatment effect: Does the treatment or intervention have real, proper, genuine, and noticeable effects on daily life?.

Perspective

This is the major point of this editorial: to emphasize clinical significance. It really does not matter if a drug is statistically better than placebo, but not clinically pertinent. It's not adequate in most conditions (except very difficult ones like intractable cancers) to get only a 52% success rate compared with say 40% on placebo: We would like to see a 90% or 97% success rate particularly with certain pain medications or antibiotics when we know that the bacteria involved are sensitive to a specific drug.

Effectively, so what if a drug is "proven" to be effective based on the statistics. We want it to work clinically and expect that. There is a role for unblemished clinical practice and that is much better than any statistical patient intervention.

We can see that double-blind studies while having their merits must be properly interpreted with their limitations. They may be a far distance away from clinical results, but should only be regarded as one component of the successful practice of medicine.

References

1. Neppe VM (2007) Double blind studies in Medicine: perfection or imperfection? Telicom 20(6): 13-23.
2. Neppe VM (1990) Ethics and informed consent for double-blind studies on the acute psychotic. Medical Psychiatric Correspondence: A Peer Reviewed Journal. Model Copy 1(1): 44-45.
3. Neppe VM (1983) Carbamazepine as adjunctive treatment in nonepileptic chronic inpatients with EEG temporal lobe abnormalities. J Clin Psychiatry 44(9): 326-331.
4. Neppe VM (1982) Carbamazepine in the psychiatric patient. Lancet 2(8293): 334.
5. Neppe VM (1981) Carbamazepine as adjunct treatment in the chronic psychiatric patient with electroencephalographic temporal lobe foci. Epilepsy International Congress, Kyoto, Japan, pp. 149.
6. Neppe VM (1989) High-dose buspirone in case of tardive dyskinesia. Lancet 2(8677): 1458.
7. Neppe VM (2014) Clinical applications of the STRAW measures of severity and frequency of involuntary movements examination in tardive dyskinesia and movement disorders.
8. Neppe VM (1999) Cry the beloved mind: a voyage of hope. Brainquest Press, Seattle, USA.
9. Neppe VM (1990) Buspirone: an anxioselective neuromodulator, in Innovative Psycho pharmacotherapy. Neppe VM (Ed.), Raven Press, New York, USA, p. 35-57.

Citation: Neppe VM (2016) Are we blind to the limits of double-blind medical studies?. J Psychol Clin Psychiatry 5(6): 00311. DOI: 10.15406/jpcpy.2016.05.00311

10. Schmeidler GR (1997) Psi-conducive experimenters and psi-permissive ones. European Journal of Parapsychology 13: 83-94.

11. Smith MD (2003) The role of the experimenter in parapsychological research. Journal of Consciousness Studies 10: 6-7.

12. Neppe VM (1982) The experimenter effect in medical research. South African Medical J 62(3): 81.

13. Rosenthal R (1963) on the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. Am Sci 51: 268-283.

14. Moss LE, Neppe VM, Drevets WC (1993) Buspirone in the treatment of tardive dyskinesia. J Clin Psychopharm 13(3): 204-209.

Citation: Neppe VM (2016) Are we blind to the limits of double-blind medical studies?. J Psychol Clin Psychiatry 5(6): 00311. DOI: 10.15406/jpcpy.2016.05.00311