

Why Double-Blind Studies?

Although most people have heard of double-blind studies, few recognize their true significance. It's not that double-blind studies are hard to understand; rather, their consequences are difficult to *accept*. Why? Because double-blind studies tell us that we can't trust our direct personal experience. This isn't easy to swallow. It's nonetheless true.

The insights provided by double-blind studies have been particularly disturbing for alternative medicine. Most alternative medicine methods are grounded in tradition, common sense, anecdote, and testimonial. On the surface, these seem like perfectly good sources of information. However, double-blind studies have shown us otherwise. We now know that a host of "confounding factors" readily create a kind of optical illusion, causing the appearance of efficacy where none in fact exists, deluding tradition and common sense, and creating meaningless anecdotes and testimonials. The double-blind study is thus much more than a requirement for *absolute* proof of efficacy (as is commonly supposed) — it is a necessity for knowing almost *anything* about whether a treatment really works. As a related issue, this fact demonstrates why it is so difficult to state with any scientific confidence which foods are healthy and which are not.

What is a Double-Blind Study?

In a randomized double-blind, placebo-controlled trial of a medical treatment, some of the participants are given the treatment, others are given fake treatment (placebo), and neither the researchers nor the participants know which is which until the study ends (they are thus both "blind"). The assignment of participants to treatment or placebo is done randomly, perhaps by flipping a coin (hence, "randomized").

Why Double-Blind Studies?

The experience of the last half century has shown that, for most types of treatments, only a randomized double-blind, placebo-controlled study can properly answer the question: "Does Treatment A benefit Condition B?"

Common sense says otherwise. It seems obvious that we can tell if a treatment works by simply trying it. Does it help me? Does it help my aunt? If so, it's effective. If not, it doesn't work.

Unfortunately, here common sense is wrong. Medical conditions are an area of life in which direct, apparently obvious observations aren't at all reliable. The insights brought to us by double-blind studies have shown medical researchers that seeing isn't believing. The reason why: a horde of confounding factors.

The Rogue's Gallery: Confounding Factors

Subtle influences called "confounding factors" can create the illusion that ineffective treatments are actually effective. It is because of these confounding factors that so many worthless medical treatments have endured for centuries. Think of the practice of "bleeding," slitting a vein to drain blood. Some of the most intelligent people

in our history were sure that bleeding was a necessity, and the medical literature of past centuries is full of testimonials to the marvelous effect of this "medical necessity."

Today, though, it's clear that bleeding is not helpful, and no doubt was responsible for killing a great many people. Why did this ridiculous treatment method survive so long? Because people were sure it worked. In reality, though, the benefits people thought they saw were really just confounding factors, such as:

The Placebo Effect

The Re-interpretation Effect

Observer Bias

Natural Course of the Illness

Regression to the Mean

The Study Effect (also called the Hawthorn Effect)

Statistical Illusions

The Placebo Effect

The *placebo effect* is the process by which the power of suggestion actually causes symptoms to improve. The original research that identified the placebo effect had some serious errors in it,¹ but there is little doubt that certain conditions are highly responsive to placebo treatment, most especially those involving pain or other subjective experience. While it's often reported that only 30% of people respond to placebo, this number has no foundation, and, in fact the response rate seen in some studies reaches as high as 70%.

Thus, if one gives a placebo treatment to a hundred people with musculoskeletal pain, one may get 70 testimonials of benefit. They will be sincere, convincing testimonials too: no one believes it's the placebo effect when it happens to them. The placebo response doesn't *feel* fake, or weak, or superficial, or shallow. It feels real.

Doctors, too, are fooled. Up until recently, orthopedic surgeons believed that "knee scraping" surgery (technically, arthroscopic surgical debridement) or knee arthritis was quite effective, and hundreds of thousands of such surgeries were performed every year. If one asked a surgeon, "how do you know this treatment works?" the surgeon would very likely reply, "Because I can see that it works with my own eyes. I have patients who go into surgery unable to walk, and a month later, they're skipping rope."

After performing this surgery for decades, one surgeon decided to use the double-blind, placebo-controlled methodology to test whether it really worked. The results were shocking: Arthroscopic surgery for knee arthritis did indeed bring about dramatic and long-lasting results; however, so did fake surgery (anesthesia and an incision), and to the same extent.² Surgeons were shocked and chagrined to find that people given the fake surgery (unbeknownst to them) were so pleased with the results that they said they would happily recommend the treatment to others!

In general, surgery lags behind other branches of conventional medicine in the extent to which it incorporates modern standards of evidence. However, there are many commonly used drugs that lack meaningful support as well. Cough syrup is a good example. In recent years, it has become clear that cough syrup containing either codeine

or the “cough suppressant” dextromethorphan doesn’t actually suppress coughs, despite decades (in the case of dextromethorphan) or centuries (in the case of codeine) of apparently effective use.³

Comparison to placebo treatment is thus essential; without such comparison, any random treatment is likely to appear effective. Few proponents of alternative medicine have grasped this basic though counterintuitive fact. It is quite common for an alternative medicine product or technique to be advocated based solely on research in which people with a problem are given a treatment, and lo and behold, they improve. But in the absence of a blinding and a placebo group, such studies are meaningless. Any nonsense treatment should be able to produce apparent improvement in a fair number of people.

Consider these examples:

In a study of 321 people with low back pain, chiropractic manipulation was quite helpful, but no more helpful than giving patients an essentially useless educational booklet on low back pain.⁴

In a randomized, controlled trial of 67 people with hip pain, acupuncture produced significant benefits, but no greater benefits than placing needles in random locations.⁵

In a double-blind, placebo-controlled study of 30 people with carpal tunnel syndrome, use of a static magnet produced dramatic and enduring benefits, but so did use of fake magnets.⁶

And in a randomized, controlled trial of 177 people with neck pain, fake laser acupuncture proved to be *more* effective than massage.⁷

Note that these studies do not actually disprove the tested therapies. The study sizes might have simply been too small to detect a modest benefit. What they do show, however, is that comparison to placebo treatment is essential: without such comparison, any random form of treatment, no matter how worthless in itself, is *likely* to appear to be effective.

Beyond the Placebo Effect

At least the placebo effect involves a real perceived benefit. Many other illusions can create the impression of benefit although no benefit has occurred at all. These are often loosely referred to using the plural term “placebo effects.”

Even when a fake treatment doesn’t actually improve symptoms, people may *re-interpret* their symptoms and experience them as less severe. For example, if someone gives you a drug and says it will make you cough less frequently, you will likely experience yourself as coughing less frequently, even if your actual rate of coughing doesn’t change. In other words, you will re-interpret your symptoms to perceive them as less severe.

Observer bias is a similar phenomenon, but it affects doctors (and researchers) rather than patients. If doctors believe that they are giving a patient an effective drug, and they interview that patient, they will observe improvements, even if there are no improvements. For a classic example of this consider the results of a study that tested the effectiveness of a new treatment regimen for multiple sclerosis by comparing it against placebo treatment.⁸ This was a double-blind study, and therefore the physicians whose job it was to evaluate the results were kept in the dark about which study participants were receiving real and which were receiving fake treatment (they were “blinded”).

However, the experimenters introduced an interesting wrinkle: they allowed a few physicians to know for certain which patients were receiving treatment (they were "unblinded").

The results were a bit appalling. The unblinded physicians were much more likely to "observe" that the treatment worked compared to the impartial blinded physicians. In other words, the unblinded physicians hallucinated a benefit because they expected to see one!

I call these results appalling because of what they say about so-called "professional experience." Suppose a physician has tried two drugs for a certain condition and found by experience that drug A is more effective than drug B. Does this mean that drug A is *actually* more effective than drug B? Not at all. If the doctor has any reason to expect drug A to produce better results (e.g., memorably positive experiences with a few patients, recommendation from a respected colleague, impressive salesmanship on the part of a pharmaceutical company), the doctor is very likely to experience drug A as more effective than drug B, regardless of the actual comparative efficacy as tested in double-blind trials. The effect is so strong that doctors are likely to discount the results of studies that don't agree with what they "know" to be true.

Many diseases will get better on their own, as part of their *natural course*. Any treatment given at the beginning of such an illness will seem to work, and a doctor using such a treatment will experience what is called the *illusion of agency*, the sense of having helped even though the outcome would have been the same regardless. A good example is neck or back pain: most episodes of these conditions go away with time, regardless of treatment, and so any treatment at all will seem to be effective.

Regression to the mean is a statistical cousin of natural course, but applies to chronic conditions. For example, consider high blood pressure. Blood pressure levels wax and wane throughout the day and from week to week. Suppose that a person's average blood pressure is 140/90, but occasionally it gets as high as 160/105 and sometimes it's as low as 120/70. If people like this are tested and found at the moment to have high blood pressure, they may be seen as needing treatment. However, if they happen to be more near their average blood pressure, or lower than their average, they won't be seen as needing treatment. In other words, when patients have a fluctuating condition, doctors (and researchers) tend to enroll them in treatment at the moment of an unhealthy extreme. By the laws of statistics, after any sufficient interval, a fluctuating variable is more likely to be measured near its mean than at its extremes. Therefore, regardless of what treatment (if any) is used, blood pressure will appear to improve with time.

The *study effect* (also called Hawthorne effect) refers to the fact that people enrolled in a study tend to take better care of themselves, and may improve for this reason, rather than any specifics of the treatment under study. This is a surprisingly powerful influence. If one enrolls someone in a trial of a new drug for reducing cholesterol, and then give them a placebo, their cholesterol levels are likely to fall significantly. Why? Presumably, they begin to eat better, exercise more, etc. Again, double-blinding and a placebo group are necessary because otherwise this confounding factor can cause the illusion of specific benefit where none exists.

Finally, *illusions* caused by the nature of *statistics* are very common. There are many kinds of these, and so they deserve a section of their own.

Statistical Illusions

Suppose you've invented a truly lousy treatment that fails almost all the time, but helps one in a hundred people. If you give such a nearly worthless treatment to 100,000 people, you'll get a thousand testimonials, and the treatment will sound great.

Suppose you give someone a treatment said to enhance mental function, and then you use twenty different methods of testing mental function. By the law of averages, improvements will be seen on *some* of these measurements, even if the treatment doesn't actually do anything. If you're a supplement manufacturer, you can selectively report the positive results to support the sales of your product, even though in fact the results are merely due to the way statistics work, and not any mind-stimulating effect of the product. (In order to validly test the mind-enhancing power of a supplement, researchers must restrict themselves to at most a couple of ways of testing benefit).

Suppose you give 1000 people a treatment to see if it prevents heart disease, and you don't find any benefit. This frustrates you, so you begin to study the data closely. Lo and behold, you discover that there is less lung cancer among people receiving the treatment. Have you made a new discovery? Possibly, but probably not. Again by the law of averages, if you allow yourself to *dredge the data* you are guaranteed to find improvements in some condition or other, simply by statistical accident.

Perhaps the trickiest statistical illusion of all relates to what are called observational studies. This is such an important topic, that again it deserves a new heading.

Observational Studies

In observational studies, researchers don't actually give people any treatment. Instead, they simply *observe* a vast number of people. For example, in the Nurse's Health Study, almost 100,000 nurses have been extensively surveyed for many years, in an attempt to find connections between various lifestyle habits and illnesses. Researchers have found, for example, that nurses who consume more fruits and vegetables have less cancer. Such a finding is often taken to indicate that fruits and vegetables prevent cancer, but this is not to be a reliable inference. Here's why:

All we know from such a study is that high intake of fruits and vegetables is *associated* with less cancer, not that it *causes* less cancer. People who eat more fruits and vegetables may have other healthy habits as well, even ones we don't know anything about, and they could be the cause of the benefit, not the fruits and vegetables.

This may sound like a purely academic issue, but it's not. Researchers looking at observational studies noticed that menopausal women who take hormone replacement therapy (HRT) have as much as 50% less heart disease than women who do not use HRT. This finding, along with a number of thoroughly logical arguments tending to show that estrogen should prevent heart disease (see footnote 23), led doctors to recommend that all menopausal women take estrogen. Even as late as 2001, many doctors continued to say that taking estrogen was the single most important way an older woman could protect her heart.

However, this was a terrible mistake. Observational studies don't show cause and effect, and it was possible that women who happened to use HRT were healthier in other ways and that it was those unknown other factors that led to lower heart disease rates, and not the HRT. Doctors pooh-poohed this objection (showing that even -- or, especially --

doctors often fail to understand the need for double-blind studies) and said that it was perfectly obvious HRT helped. However, when a double-blind, placebo-controlled study was done to verify what everyone “knew” was true, it turned out that that HRT actually *causes* heart disease, rather than prevents it.⁹ HRT also increases risk of breast cancer, and, possibly, of Alzheimer’s disease as well. In other words, placing trust in observational studies may have led to the deaths of many women. This is not an academic issue.

In hindsight, it appears that women who happened to use HRT were healthier not because of the HRT, but because they tended to be in a higher socioeconomic class, have better access to healthcare and also take care of themselves. (From this perspective, women taking HRT in observational studies had less heart disease in spite of taking HRT, not because of it.) However, it is also possible that the real cause of the spurious association between HRT use and reduced heart disease rates is some other factor that we have not even identified. People’s characteristics are not independent variables; they are linked in numerous complex ways. For this reason, observational studies can’t prove cause and effect, and they can easily lead to conclusions that are exactly backwards.

This is a lesson that the news media seems unable to understand. It constantly reports the results of observational studies as proof of cause and effect. For example, it has been observed that people who consume a moderate amount of alcohol have less heart disease than those who consume either no alcohol or too much alcohol. But, contrary to what you may have heard, this doesn’t mean that alcohol prevents heart disease! It is likely that people who are moderate in their alcohol consumption are different in a variety of ways from people who are either teetotalers or abusers, and it may be those differences, and not the alcohol per se, that causes the benefit. Maybe, for example, they are moderate in general, and that makes them healthier. The fact is, we don’t know.

Similarly, it has been observed that people who consume a diet high in antioxidants have less cancer and heart disease. However, once more this does NOT mean that antioxidants prevent heart disease and cancer. In fact, when the antioxidants vitamin E and beta-carotene were studied in gigantic double-blind studies as possible cancer- or heart-disease-preventive treatments, vitamin E didn’t work (except, possibly, for prostate cancer) and beta-carotene actually made things worse! (One can pick holes in these studies, and proponents of antioxidants frequently do, but the fact is that we still lack direct double-blind evidence to indicate that antioxidants truly provide *any* of the benefits claimed for them. The only evidence that does exist is directly analogous to that which falsely “proved” that HRT prevents heart disease!)

Double-Blind Studies, and Nothing but Double-Blind Studies

All of the information I’ve just presented has accumulated over the last several decades. After coming to a great many false conclusions based on other forms of research, medical researchers have finally come to realize that without doing double-blind studies on a treatment it’s generally impossible to know whether it works. It doesn’t matter if the treatment has a long history of traditional use -- in medicine, tradition is often dead wrong. It doesn’t matter if doctors or patients think it works -- doctors and patients are almost sure to observe benefits even if the treatment used is fake. And it doesn’t matter if observational trials show that people who do X have less of Y. Guesses

made on the basis of this kind of bad evidence may be worse than useless: they may actually cause harm rather than benefit.

To make matters even more difficult, double-blind studies are not all created alike. There are a number of pitfalls in designing, performing and reporting such studies, and for this reason some double-blind studies deserve more credence than others. Double-blind studies from certain countries, such as China and Russia, always must be taken with a grain of salt because historical evidence suggests a pattern of systematic bias in those countries.¹⁰ Studies that enroll few people, or last for only a short time, generally prove little. And unless more than one independent laboratories have found corroborating results, there's always the chance of bias or outright fraud. Thus, a treatment can only be considered proven effective when there have been several double-blind studies enrolling 200 or more people, performed by separate researchers, conducted according to the highest standards (as measured by a study rating scale called the "Jadad scale"), carried out at respected institutions and published in peer-reviewed journals. Weaker evidence provides, at best, a hint of effectiveness, likely to be disproved when better studies are done.

While a number of herbs and supplements have reached, or nearly reached the level of solid proof, most alternative therapies have not.¹¹ Again, this isn't an idealistic, ivory-tower standard useful only for academia: it's a minimum necessity. Treatments that have not been evaluated in double-blind studies are very likely to be so much hot air. Except in the rare cases when a treatment is overwhelmingly and almost instantly effective (a so-called "high effect-size" treatment), there is simply no other way to know whether it works at all besides going through the trouble and expense of double-blind trials.

Evidence-Based Medicine

The double-blind study has begun to cause a revolution even in conventional medicine. Many old beliefs have been tossed out when double-blind studies were finally done. It's been discovered, for example, that (as noted earlier) over-the-counter cough syrups don't work, that immediate antibiotic treatment for ear infections is probably not necessary or even helpful in most cases¹² (the same goes for sinus infections), and that cartilage scraping for knee arthritis is no better than placebo (but, as noted above, placebo is very effective!). Many other widely used conventional treatments, such as spinal fusion for back pain and most physical therapy modalities for any sort of pain have essentially no supporting evidence, even if they haven't yet been *proven* ineffective. Regarding back fusion surgery, at least, this is a big problem; unproven surgery borders on the criminal.

The understanding that medicine must be grounded in double-blind studies is called the "evidence-based medicine" movement. According to evidence-based medicine, if a treatment has not been properly studied, it should not be advocated as an effective treatment.

This is true whether the treatment is an Indonesian herb or a well-accepted medical technique. However, conventional medicine, at least, has a certain reticence about offering unproven treatments, and a historical trend of rising evidence standards. Alternative medicine has taken the opposite approach: offering a profusion of treatments without the slightest concern about meaningful scientific support.

Double-blind studies on alternative medicine are at last coming in. Unfortunately, the results aren't all that great. Alternative medicine contains a few modestly effective individual treatments, but its grand schemes of healing most likely possess no more truth than other pre-scientific approaches to understanding the physical universe. The body is a biochemical machine, alas, not a spiritual entity composed of subtle energies, and one can't get very far without a microscope.

Perhaps even more disturbing, the necessity of double-blind studies shows why we know so little about which foods are healthy and which are not: there is no way to do a double-blind study of foods. (How would participants in such a study fail to know what they were eating?) Because of this, the best one can usually come up with for evaluation of dietary approaches are the same tired, weak, fallible, untrustworthy observational studies. This lack of evidence in an area where people want guidance leads to a profusion of imaginary dietary advice, and, hence, to the underpinnings of orthorexia.

¹ Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med.* 2001;344:1594-1602

² Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med.* 2002;347:81-8.

³ See: Taylor JA, Novack AH, Almquist JR, et al. Efficacy of cough suppressants in children. *J Pediatr.* 1993;122:799–802; Croughan-Minihane MS, Petitti DB, Rodnick JE. Clinical trial examining effectiveness of three cough syrups. *J Am Board Fam Pract.* 1993;6:109–115. Schroeder K, Fahey T. Over-the-counter medications for acute cough in children and adults in ambulatory settings. *Cochrane Database Syst Rev.* 2001;CD001831; and, most impressively, Paul IM, Yoder KE, Crowell KR et al. Effect of dextromethorphan, diphenhydramine, and placebo on nocturnal cough and sleep quality for coughing children and their parents. *Pediatrics.* 2004;114:e85-90.

⁴Cherkin DC, Deyo RA, Battie M, et al. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. *N Engl J Med.* 1998;339:1021–1029.

⁵Fink M, Karst M, Wippermann B, et al. Non-specific effects of traditional Chinese acupuncture in osteoarthritis of the hip: a randomized controlled trial. *Complement Ther Med.* 2001;9:82–88.

⁶ Carter R, Hall T, Aspy CB, et al. Effectiveness of magnet therapy for treatment of wrist pain attributed to carpal tunnel syndrome. *J Fam Pract.* 2002;51:38–40.

However, identical benefits were seen among those given fake magnets.

⁷ Irnich D, Behrens N, Molzen H, et al. Randomised trial of acupuncture compared with conventional massage and sham laser acupuncture for treatment of chronic neck pain. *BMJ.* 2001;322:1–6.

⁸ Noseworthy J H, Ebers G C, Vandervoort M K, et al. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology.* 2001;57:S31-5.

⁹ Manson JE, Hsia J, Johnson KC, et al. Women's Health Initiative Investigators. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med.* 2003;349:523-34.

¹⁰ Vickers A, Goyal N, Harland R, et al. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials.* 1998;19:159-166

¹¹ To be fair, for some types of treatment, such as chiropractic, acupuncture, physical therapy and surgery, it isn't possible to design a true double-blind study: the practitioner will inevitably know whether real or a fake treatment has been applied. In such cases, most researchers settle for a "single-blind" design, in which the study participants (and the people who evaluate the participants to see if they've responded to therapy) are kept in the dark, but not the practitioners of the therapy. The problem with such single-blind studies, though, is that the practitioners may convey enthusiasm when they are providing a real treatment and lack of enthusiasm when they apply a fake one.

The former might act as a better placebo than the latter, and thereby produce the results that really have nothing to do with the treatment itself. To get around this, Kerry Kamer D.O. has suggested using actors trained to provide fake treatment with confidence and enthusiasm, but, so far as I know, this has not yet been tried.

¹² Pappas DE, Owen Hendley J. Otitis media. A scholarly review of the evidence. *Minerva Pediatr.* 2003;55:407-14.