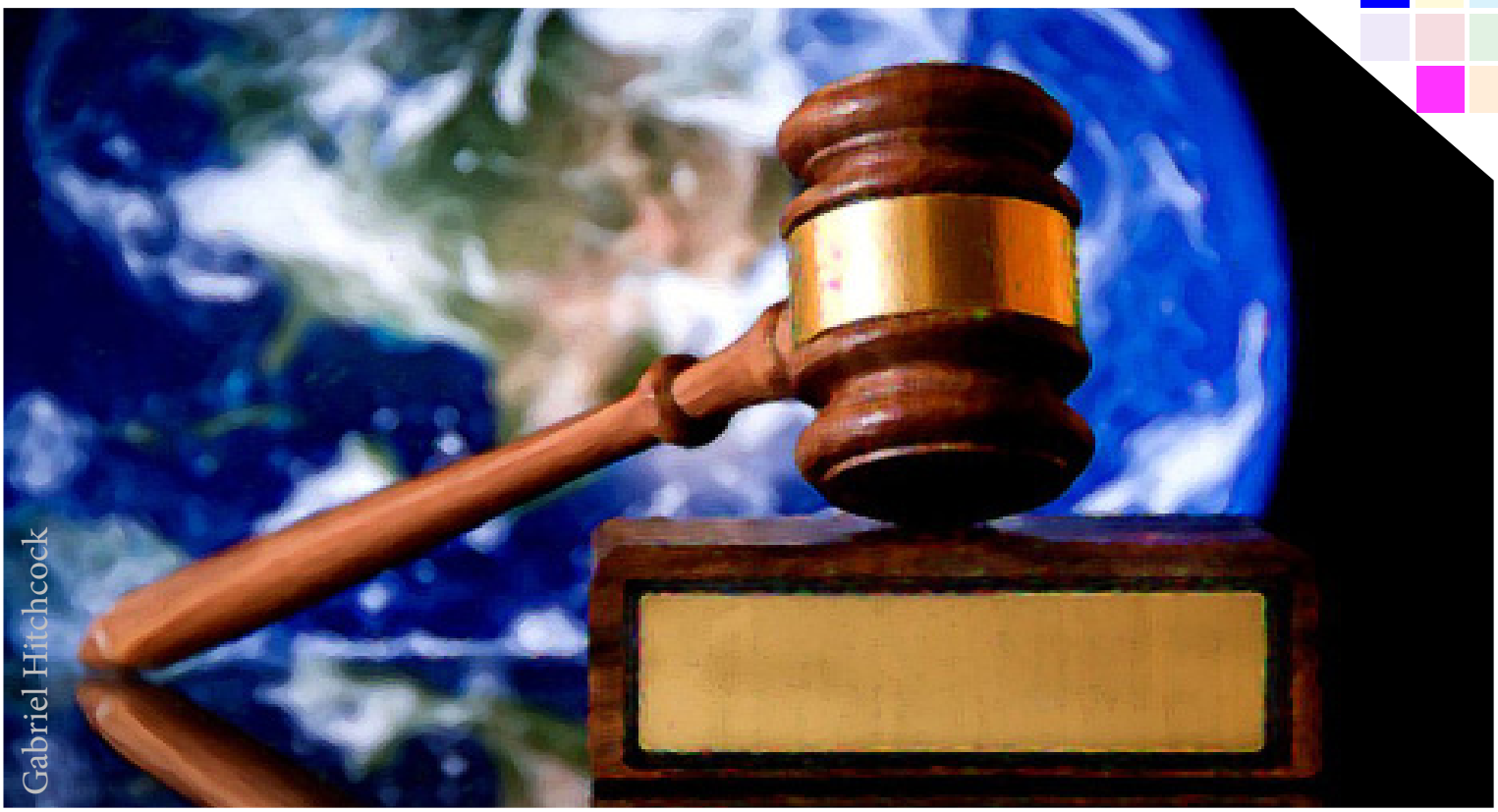
The Synapse: Intercollegiate science magazine

Volume 5 | Issue 1 Article 8

2015

Your Treatment on Trial

Kevin Ng



Your Treatment on Trial

Kevin Ng

How do you know whether a medical treatment works or not? Well, there are a number of ways you could find out. You could try it out for yourself and see if you get better afterward, or you could see how a friend fares after your friend gives it a try. Unfortunately, there isn't a way to tell the difference between improvement due to the actual treatment, improvement due to the placebo effect, improvement due to the natural course of your symptoms, etc. It all feels the same to the person experiencing the improvement in condition. Another way you could find out is ask your physician. Unfortunately, physicians cannot distinguish between the different causes of improvement either; patient improvement is patient improvement, to the watchful doctor. Also, a doctor might prefer treatment A over treatment B, not because of superior evidence, but maybe because of tradition (think of bloodletting), or because the doctor's team of colleagues all have had good experiences with treatment A, or because the doctor is selectively remembering the times treatment A worked and unknowingly forgetting the patients for whom it didn't.

Small samples of patients, unconscious cognitive biases, inability to distinguish between different causes, etc. all get in the way of evaluating the true effectiveness and safety of a treatment. So now what? We all have only our own experiences to go off of; we can't go back in time, switch our treatment for a placebo at the last minute, and see if we get the same result. Without any comparison of experiences, there is no way to consider alternative explanations for an effect. Well, instead of looking at an individual experience with the treatment, we can collect a whole bunch of experiences; we can look at those bunches of experiences in the context of each other. When we randomly assign people to get different treatment experiences, we have a trial. In medicine, a trial compares a treatment against another treatment, or a placebo, or a wait-list control.

Why randomly assign? If researchers were allowed to pick who went in what group, they could, either consciously or unconsciously, assign people to the groups such that one group is different from the other in a way that biases the trial. If all the young, healthy people with fewer medical problems were in the treatment group, and all the old people who smoke and have multiple medical issues were in the control group, then you can make the treatment look really good since the people in the treatment group would be more likely to improve anyway. But suppose

people who exercise a lot are just as likely to get into the treatment group as the control group. Same for people who don't exercise. Moreover, people who recently started a fad diet are just as likely to get into the treatment group as the control group. Same for people who did not recently start a fad diet. The rate at which people get better on their own would also be about the same for both. Headaches come and they go, back pain feels like the worst thing in the world at some points in time and not so bad at others, and the severity of your stuffy nose isn't the same 24/7 throughout your week of cold. These fluctuations happen even if you don't do anything. The factors that affect the ebb and flow of your illness would be randomly distributed. The point is, random assignment assures that the groups are roughly the same in composition, dispersing the factors we don't want to influence the results, and isolating what we care about: the effects of being given the treatment.

A lot of the time, the control group is a placebo group. Basically, a placebo is a supposedly inert form of the treatment, and the placebo effect refers to therapeutic improvement brought about by one's beliefs and expectations. A lot of studies have been done on the placebo effect. It seems that there's a dose response; taking four sugar pills, with no active ingredient, is a more effective treatment for pain relief than taking two sugar pills. Likewise, a salt water injection brings about more relief than sugar pills. Sham surgery, sham acupuncture, and other sorts of "pretend" treatments have been shown to elicit a therapeutic response. Of course, these are average effects; the exact response will vary based on the particular patient's expectations.

Ideally, the trial is double-blinded, meaning neither the patients nor the physicians know who is in what group. This is to prevent biasing that would influence the outcome. If a patient knows he or she is in the placebo group, the patient will think, "Hey, I'm in the placebo group. This is a joke; I'm not getting anything – I'm not going to get better." If the patients in the treatment group know so, they'll think, "Wow, I'm getting the real deal here. I'm totally going to get better." Obviously, this would affect their expectations, enhancing the placebo component of the actual treatment response and downgrading the placebo group's placebo responses. And the doctors? If the doctor knows who is in what group, these expectations can change the doctor's conduct and tone of voice toward the patients in the different groups, unconsciously breaking the blinding. This means it might be unintentionally communicated to the placebo people that they are in the placebo group and to the treatment people that they are in the treatment group. Another effect of this knowledge is interpretation. Suppose you think that women draw better circles than men, and you run an experiment to find out by recruiting a bunch of men and women. Each person writes his or her name at the top of a sheet of paper and then draws a circle. You collect the stack of papers and judge the roundness of the circles on each sheet. Now, judging the roundness of a circle, like observing clinical features in a physical examination, is not a black-and-white endeavor. There is ambiguity, and subjective judgment calls must be made. The more subjectivity is involved, the more room there is for those unconscious biases to creep into your decision-making process. If you notice the name at the top of the paper looks female-typical, you might be more likely to rate the circle as being rounder than not; if the name looks male-typical, you might rate the circle as being less than perfect. All in all, your judgment calls would have been skewed by your knowledge and prior beliefs. It would have been better to conceal the names beforehand. You can imagine how knowing who is getting the placebo and who is not would play out on a doctor's expectations, and in turn the doctor's impression of who is getting better and who is not.

Once all is said and done, you look to see if more people in the treatment group got better than the control group, or if they got better faster, or if their side effects were not as bad, or if fewer people died, etc. Random assignment, the placebo effect, and double-blinding are essentially three uber important qualities of a well-done Randomized Clinical Trial, or RCT. There are other aspects that are important for how a RCT is conducted and analyzed, such as representativeness of patients, sample size, surrogate outcomes, taking account of dropouts, external validity, primary vs. secondary outcomes, etc. However, the basic skeleton of a quality RCT comprises these three elements. When one or more of these are missing, it detracts from the reliability of the RCT as a source of evidence for the efficacy of the treatment. And when you gather a bunch of RCTs in a systematic review to look at the big picture, those detractions add up and can make the review less than ideal. After all, junk in means junk out; synthesizing a bunch of flawed studies does not reduce the flaws.

What's not to like about RCTs? Well, RCTs cannot tell you anything about the mechanism of action. Maybe the treatment works by speeding up the work of particular enzymes; maybe the individualized nature of the treatment is part of the mechanism of its effectiveness; maybe it works through some yet to be discovered process or substance. When the first RCT was done on scurvy in ships on the sea, it was found that lemons did a really good

job at reducing scurvy related death. At the time, nobody knew about the existence of Vitamin C, or how Vitamin C deficiency caused scurvy, yet the results of the trial were astonishing. The treatment worked; sailors eating lemons fully recovered, compared to the sailors drinking vinegar or doing nothing, who were still suffering. The details about how it worked could be worked out later.

Of course, it helps to be able to conceive of a mechanism since it means there's plausibility. In fact, if a treatment's implied way of working violated the Second Law of Thermodynamics, the plausibility of the treatment would be so low that spending the resources to do an RCT and using up patients' time would be inappropriate. Even if a positive relationship were found, it should be met with skepticism since it would likely be a false positive, or chance occurrence resulting from random noise in the data. There is a reason why scientists are not taking the time to test the hypothesis that giving buttered toast to kittens gives them telekinetic powers, or if holding an anvil over your head as you jump out of a plane will reduce your likelihood of dying.

Other criticisms of RCTs have to do with medicine itself. Some think that group statistics do not apply to individuals, or that relying on RCTs instead of clinical experience will result in treating patients uniformly and coldly as numbers. The former is a misunderstanding; the unique characteristics that make each of us an individual do not necessarily undermine the effectiveness of an intervention. Those unique variables might be irrelevant to the underlying mechanism of the treatment. In other words, the unique attributes of each individual may not interact with the intervention, or they may be overcome by its main effects. Age, sex, ethnicity, other medical conditions, etc. may or may not affect the effectiveness or safety of the treatment. We would have to have done the RCTs to begin with and notice variation among subgroups, suggesting further research in particular groups of people. Here's an example: each person with melanoma is a unique individual – they vary in age, sex, hair color, diet, lifestyles, and who knows what else. But this doesn't change the fact that 90% or more of cases of this form of skin cancer are mostly curable with early surgery.

The second criticism is not so much of a problem if we put RCTs in context. RCTs play a role in clinical-decision making, and so do clinical judgment and patient values and preferences. They answer questions like "Is drug A better than drug B at reducing the risk of death from heart disease?" or "How does eating tofu affect your probability of recovering from stroke?" When we want to find out whether a treatment works or not, RCTs make up for the flaws of clinical judgment. Physicians are never blinded; they, like the rest of us humans, can exhibit cognitive errors that skew how they think; they might rely on what they learned decades ago; they might just be "going with the flow" by doing what's popular; they might be convinced by a pharmaceutical representative's spiel. They extrapolate from the results of previous clinical cases in order to figure out what to do with the next patient. The difference between that and the RCT database is that the research literature comprises a much larger sample of asthma patients, cancer patients, depression patients, etc. than what any particular clinician encounters in his or her clinical career.

Yet, clinical experience is valuable. In skill-based technical procedures, like surgery, experience is what fine tunes your abilities. Less tangibly, there are still many, many questions for which there has yet to be quality research done. Clinical expertise tells a doctor what to do in the absence of evidence, and this is where the art of medicine comes in. If a patient would forgo a more effective treatment in order to get fewer side effects, the doctor takes that into account. If a patient would prefer a treatment with more side effects if it's more effective, the doctor takes that into account. If a patient would prefer doing nothing, and seeing how he or she turns out, the doctor takes that into account. If the patient's cultural background limits what options are available, the doctor takes that into account. RCTs cannot tell the doctor what the patient ultimately wants – only the physician-patient relationship can. If RCTs are used as the final arbiter of clinical decisions, without listening to the patient, then this is cookbook medicine – it treats patients as numbers on a paper and not as suffering human beings who would like some care.

RCTs are not done for that purpose; they are a valuable source of evidence that plays a part in the clinical decision-making process. Efficacy and safety data do not, and cannot, replace a strong physician-patient relationship. The data can only inform, and RCT data should be put in the context of the other data – basic science lab studies, observational studies, case reports, etc. Each source of evidence has its pros and cons, and looking at any single category of evidence without context is useless. The whole body of evidence is what matters. RCTs are not perfect, but they are awesome enough; let's not overdo it and push them to do what they were not designed to do.