

# somatic research

RAVENSARA S. TRAVILLIAN



---

# EVALUATING ACCURACY OF A HYPOTHESIS

**What scientific studies do is test the null hypothesis—that is, they examine whether it is true that a given treatment makes no difference in the situation at hand.**



In the previous column, we shifted from a high-level view of the IMRaD structure of research articles to begin looking more closely at how the introduction sections of articles are put together—especially as needed to state a clear research hypothesis. Now let's dig a little deeper into what evaluating a research hypothesis means. We'll continue to use some instances of pregnancy massage research to illustrate the main points.

## Check Your Math

**Remember,  $p$  represents a probability that—if the null hypothesis were actually true—that we would see the results we see from the study. A probability of 1 means that an event is certain—it will happen with 100% certainty. Similarly, a probability of 0 means that an event will absolutely never happen—it is certain not to happen, or there is 0% chance of it happening. So the  $p$  values you will see in the literature will always be between 1 and 0, never anything else: if you ever do see a larger or smaller  $p$  value, it's a mistake. Similarly, when you convert the decimal representation of  $p$  to a percentage, the answer you get will always be between 100% and 0%. If you get an answer outside this range, go back and check to make sure you moved the decimal point exactly two places to the right, no more, no less.**



Remember, a hypothesis, or research assumption, describes relationships among things in the world. For example, the hypothesis “massage increases relaxation” makes a claim: when massage treatment is applied, relaxation is expected to increase.

Some of those descriptions depend on being precise about what's what: which kind of massage are we talking about? For how long? How do we measure relaxation? And so on. Those are big questions, and we'll delve into them in later columns. Here we'll focus on a widely-used statistical measure that judges whether a proposed hypothesis is correct or incorrect. We'll briefly sketch the basics of what you need to evaluate the statistical results you'll see when you look at research articles independently.

### THINKING STATISTICALLY: SIGNIFICANCE, PART II

As we know, reality is not a vending machine where we put in a coin and a clear, indisputable fact pops out. There would really be nothing to research (and it would be pretty boring) if living things were so simple and obvious as that. Cause-and-effect linkages are very complex, can appear to be different than they really are, and sometimes lead us into making errors. For example, we might think that massage has an effect on a particular patient's recovery from an injury, when the real effect came from the body's natural abilities to heal. That's an example of a false positive: thinking there was a cause-and-effect relationship when there really wasn't. The opposite error can happen, too: some people think massage “doesn't do anything,” a false negative—when

in reality much evidence shows clear positive effects for massage. That's an example of missing a cause-and-effect relationship that really is there.

Like the two kinds of errors, there are also two ways of being right about observing cause and effect. The first way is to correctly detect an effect that actually exists. For example, if a client has pain, and massage really does relieve the pain, then we have made a correct observation that massage relieved the client's pain.

Similarly, if there is no cause-and-effect relationship, and we correctly observe no cause-and-effect relationship, that is also a way of observing reality correctly. For example, if a hospital patient in the intensive care unit requests a massage to relax and feel less anxious, the medical staff may be hesitant to provide it if they know that massage can reduce heart rate in a resting patient and are concerned that massage may cause the patient's heart rate to fall dangerously low. If massage does not have that effect in reality, as demonstrated by studies that show massage has no dangerous effect on heart rate, that's an example of the second way an observation can be right.

	Case: There is a real effect.	Case: There is no real effect.
Observation: We detect an effect.	Match: True positive.	No match: False positive.
Observation: We detect no effect.	No match: False negative.	Match: True negative.

Table 1. There are two ways of being right when the observation matches the case in reality, and there are two ways of being wrong when they do not match.

The null hypothesis is a way to deal with such complications. Since a hypothesis describes relationships among things in the world, the null hypothesis represents the assumption that the difference between the treatment (massage) and control (comparison group which receives no massage) groups is null, or nothing: that there is no difference at all. That would mean, in turn, that there is no relationship between the things described in the hypothesis.

What scientific studies do is test the null hypothesis—that is, they examine whether it is true that a given treatment makes no difference in the situation at hand. The null hypothesis is that there is no difference between treatment and control groups. So if, with a reasonable amount of certainty, we show that the null hypothesis is not true, it means that there is a difference between the treatment and control groups: the treatment effect. If we have done a good job of designing the research study, and the groups are like each other in all ways except for the fact of receiving or not receiving a massage, then—presumably, provisionally—the difference between the two groups may be attributable to the massage.

A way of determining statistical significance is to determine the probability of getting the results generated by the study, if the null hypothesis is true. Remember, if we see a difference, it should mean that the null hypothesis is false. However, just because of chance, sometimes we will see a difference, even if the difference is not there. Researchers design studies to make this possibility as low as they possibly can, but for all practical purposes, it can never be zero.

“Statistically significant at the .05 level,” or similar wording, is something you will come across a lot in the literature. It means that the researcher has decided to assign statistical significance to a variable  $p$ , which represents probability if the reported value of  $p$  is 0.05 (5%) or less, then the observed effect is considered to have statistical significance. If, on the other hand, the reported value of  $p$  is more than 0.05 (5%), the observed effect is not considered to have statistical significance. For  $p$ , at any rate, bigger is not better when it comes to statistical significance. There’s nothing magical about the 0.05 number chosen—other researchers may choose other numbers for other studies—but this value is a popular choice, for reasons which will become clear as we step through it.

Let’s take a closer look at the following:  $p < 0.05$  (5%). This means the probability that the null hypothesis is true, and there were no real changes between the massage group and the

non-massage group—that the study results were due to chance or some other factor than the treatment—is less than 5%. So if we have made an error, and we think that massage is having an effect when it really isn’t (a false positive), we can expect to make that error fewer than five times out of every hundred that we perform the same study.

That’s why a lot of researchers like a  $p$  of 0.05 or less—it’s an intuitive benchmark that’s easy to understand. While not a perfect rate—it leaves one time out of 20 when we could expect to have an error due to chance—that rate is still a very good one, good enough so that many researchers will aim for it, if the hypothesis and the research design allow. Many articles in the research literature display results that are valid (another concept we’ll discuss later) at the 0.05 (5%)  $p$ -value.

What this value for statistical significance means is that it is very unlikely that we are seeing an effect in this study that is not really there. It is much more likely that the effect we are seeing is a real one. So we reject the null hypothesis—and provisionally, subject to later results from further studies, we accept the study results.

## READING RESULTS

Let’s take another example through that process. This  $p$ -value is from a study where women who reached the 40th week of pregnancy (the expected delivery date range) without going into labor were taught shiatsu techniques by their midwife. The hope was that shiatsu would help them progress into labor spontaneously, on their own, so that they would not have labor induced by medical or pharmacological means.



SEE THE BLOG AT [HTTP://SOMATIC-RESEARCH.BLOGSPOT.COM](http://somatic-research.blogspot.com) FOR SOME PREGNANCY MASSAGE EXAMPLES TO PRACTICE ON. NEXT TIME: THE "METHODS" SECTION, STUDY SIZE AND POWER, AND SPORTS MASSAGE.

Cause-and-effect linkages are very complex, can appear to be different than they really are, and sometimes lead us into making errors.



*Results: Post-term women who used shiatsu were significantly more likely to labor spontaneously than those who did not ( $p = 0.038$ ). Ingram J, Domagala C, Yates S. The effects of shiatsu on post-term pregnancy. Complement Ther Med. 13, no. 1 (2005): 11–5.*


We start with the p-value reported by the researchers,  $p = 0.038$ , and multiply it by 100 to arrive at its corresponding percentage:  $0.038 \times 100 = 3.8\%$ .

We have the new value “3.8%,” and we plug that new value back into the p-value statement we started with:  $p = 3.8\%$ .

Again, we have a p-value less than 5%, and so the observation that shiatsu aided the pregnant women in spontaneously going into labor, rather than needing to have it brought on, has statistical significance in this study.

### USING STATISTICAL SIGNIFICANCE IN DEVELOPING RESEARCH LITERACY

In the complex and diverse reality of life, where different people react differently to similar things, and relationships between causes and effects are not simple and linear, we can correctly distinguish true positives and true negatives, or sometimes we can fall into false positive or false negative errors. Statistical significance is one tool that we can use to orient ourselves about what is really going on in a study, and to distinguish between the two ways of being right, versus the two ways of being mistaken about cause and effect. Reading that a particular study has statistical significance at, for example, the 0.05 (5%) level, indicates that we are sufficiently confident that we are seeing a real cause-and-effect connection between the treatment and the outcome in that study, that we would reasonably expect to make one of those errors only 1 time out of every 20 repeats of the experiment, or only 5% of the time. **m&b**

 Ravensara S. Travillian is a massage practitioner and biomedical informatician in Seattle, Washington. She has practiced massage at the former Refugee Clinic at Harborview Medical Center and in private practice. In addition to teaching research methods in massage since 1996, she is the author of an upcoming book on research literacy in massage. Contact her at [researching.message@gmail.com](mailto:researching.message@gmail.com) with questions and comments.