


somatic research

BY RAVENSARA S. TRAVILLIAN



THE METHODS SECTION

Of all the sections of a research article (Introduction, Methods, Results, and Discussion), the Methods section is the most critically important to the validity of the study's outcome. 

In the previous column, we examined the null hypothesis—a technique used by researchers to determine what the evidence reveals about their own hypothesis, or research question. In the studies we're most interested in, the null hypothesis often states something to the effect that there is no difference in positive outcome between the treatment group who received massage and the control group who did not receive massage.

If the evidence provides a reason to do so, then researchers reject the null hypothesis: they determine that it is indeed false to say that there is no difference between the two groups. That means we have a certain degree of confidence in provisionally and contingently accepting the research hypothesis that massage did make a difference—that is, in accepting it until and unless further evidence indicates that the question needs to

be revisited. In this column, we'll talk more about what that means for us in reading the massage research literature, and we'll take some examples from sports massage literature to illustrate the main points.

THE METHODS SECTION

The Methods section (the M in the IMRaD mnemonic for the structure of a research article) is where the researchers lay out for the reader how they plan to test the research question they introduced in the Introduction section. There, you'll find not only the description of the massage treatment being tested, but also a section of how data will be collected and measured and what statistical and other tests will be used to evaluate the data.

Of all the sections of a research article (Introduction, Methods, Results, and Discussion), the Methods section is the most critically important to the validity of the study's outcome. Valid results can always be reinterpreted in light of further knowledge; discussions of the study's meaning for practice can always be expanded upon later—but if the study itself is flawed, then the results that come out of those flawed methods will themselves be less trustworthy. Since trust in the groundwork laid by the researchers who went before is crucial to any scientific endeavor, results of a study have to be as strong as possible to support the studies they in turn will lead to, and the practitioners who will rely on those results. The methods used are crucial to ensuring that the study results are as valid as possible. What this means in practical terms is that the researchers will discuss what steps

they took to strengthen their methods and what challenges remain to consider in evaluating the study. Those steps include such things as making sure the study was carried out on a large enough population—a sufficient sample size—to be as sure as possible that they are seeing the results they think they see. As the reader, your job is to evaluate how well they succeeded at their task.

The ideal, of course, is a perfect methodology, but in the real world, researchers have to work around methodological issues to make the results as dependable as possible. Not only are there funding constraints and other pragmatic and logistical issues that work against the design of a perfect study, but there are also trade-offs between study issues that have an impact on methodology.

For example, if you want your massage research protocol (your “treatment recipe”) to be repeatable by other researchers, you have to spell it out in detail for them to follow. But if you do that in advance, then you’re shutting out the interactive part of massage, where the therapist responds to the verbal and nonverbal feedback from the client—so the massage protocol is not very representative of what a session is really like. You could, of course, free up the therapists in your study to do whatever they would normally do in a session, responding freely and interactively to the needs of the client in the moment, but then, how would you tell researchers who want to replicate your study later how they should proceed? Those two important concerns, real-world representativeness and replication by other researchers, are actually in opposition to each other, because as one increases, the other necessarily decreases.

So even in theory, assuming unlimited funding and total availability of other resources, methodological perfection is an unattainable goal. The best that researchers can do instead is a balancing act—make the study

The risks of false negative and false positive errors can never be totally eliminated.



as strong as possible and give the reader a heads-up on what factors of the study design should be taken into account in interpreting the results of the study. So even in good and strong research articles, you will find the researchers addressing “weaknesses” or “limitations” of the study.

How all of these factors strengthen or weaken the study will vary in different contexts. For example, researchers studying a hospital-based treatment may focus on a very strict protocol to achieve the replicable and quantitative results they want in their particular situation, while researchers studying parents massaging ill children in their own rooms may focus less on the protocol and more on the interaction between the parents and children. Considerations like this make methodological issues and trade-offs into open-ended questions, so there is no checklist to use for every article. Think about what the ultimate purpose of the research is and how well researchers succeeded at their stated goals.

POWER AND SAMPLE SIZE

One limitation often found in massage research methods relates to study size—you’ll find statements in the literature like, “Most studies contain methodological limitations including ... few subjects ...”,¹ or “These conclusions are limited by the small sample size of the included [research studies].”² Clearly, when it comes to results, something methodologically important is going on with small studies. Additionally, you may have heard people say a massage research study needs about 35 or 40 people, more or less, to have a large enough sample size—what’s up with that? What’s so special about that number?

Like the indicator of statistical significance p discussed in the last issue, the power of a test is a probability. In this case, it is the probability that the test will not make a Type II error (false negative) by missing a treatment effect that is really there. When $p = 0.05$, for example, it represents a 5% chance, or 1 time out of every 20 that you rerun the study, that you would make a Type I error (false positive), or think that you were observing a real effect, when it was really due to chance.

While there is no universal measure of power, you’ll often see 0.80 as a target that researchers aim for—it means that they expect that 80% of the time, or 4 times out of 5, if there is a treatment effect in the study, they will detect it. (Remember, for both p and power, when it is represented as a decimal number, multiply that number by 100 to get the percentage it represents.) The risks of false negative and false positive errors can never be totally eliminated, but judicious use of statistical significance and of power allow both of those risks to be managed, resulting in a certain degree of confidence in the validity of the study results.

The ideas of statistical power, sample size, and the null hypothesis are tightly linked to each other. For reasons we’ll get deeper into in a



later discussion, researchers look at the evidence to see whether it calls for rejecting the null hypothesis and supporting their own hypothesis. For example, if a researcher hypothesizes, like Jönhagen's ("Sports Massage After Eccentric Exercise") team did, that "Sports massage can improve the recovery after eccentric exercise,"³ then the null hypothesis would be something like "Sports massage has no effect on recovery after eccentric exercise." All of these concepts come back, ultimately, to whether to accept or reject the null hypothesis.

As it happens, Jönhagen's team did end up accepting the null hypothesis and rejecting his research hypothesis, because they found that the massage

had no effect on their measurements of quadriceps pain, strength, or function after the exercise. We'll get back to the larger implications of those findings toward the end of the article, but here, we'll just talk about the null hypothesis. A goal of a research study is to try to correctly determine whether or not to accept or reject the null hypothesis—neither to accept it mistakenly (false negative) nor to reject it mistakenly (false positive). To see how that works in practice, we'll switch from sports massage to cardiac surgery for a moment, since a particular research article demonstrates clearly how the researchers calculated a power analysis for their study.

Hattan's ("The Impact of Foot Massage and Guided Relaxation Following Cardiac Surgery: A Randomized Controlled Trial") research team investigated whether foot massage and guided relaxation promoted calmness (among other measures) in cardiac surgery patients. Their description of how they determined the ideal sample size for their study points at the multiple factors involved: "A post hoc [carried out after the study] power analysis test suggested that a sample size of 45 would be required to detect a difference of the size observed with an acceptable level of Type II error [false negative] (power = 0.8)."⁴ From this statement, we can see that statistical power has to do with detecting an effect, with the size of a sample, and with how much risk of error we're willing to tolerate. In the literature, you'll often see it written in a much shorter way, but Hattan's description shows details of what is involved in a power analysis—sample size, effect size, and acceptable tolerance of error.

One way to think of it is, how large a population do you need to make sure you see an effect that is there—that you don't make a false negative error by missing something? If it's a large effect, you probably don't need as

many people to see it as you do if it's a small effect—in other words, if it's something that could be easily missed, you improve your chances of seeing it by looking for it in more people. But if it's a major effect, it will probably show up more dramatically, and you can see it in fewer people. For that reason, increasing sample size is a very common way of increasing the power of a test.

So where did that often-mentioned number 35–40 for massage studies referred to earlier come from? It's an estimate that probably came out of one particular study as having sufficient power in that context, and was then accidentally generalized into a more universal number that is sometimes quoted as applying to many massage research studies. But since a sufficiently large sample size depends on the size of the effect being looked for, and how much risk of error the researchers are willing to accept, it really depends on the question being researched. When researchers design a study, they put a lot of time and effort into the question of how many participants to include, and they consult statisticians to determine that number, because they know that funding agencies will examine it carefully to determine whether they've gotten it right.

There's no "one size fits all" number that massage research studies should have to ensure sufficient power. Instead of trying to come up with such a number for all studies, a better strategy is to follow the researcher's logic, as explained in the article, for why that particular number was right—ensured sufficient power—for that study on its own terms. If the researchers' explanation of how the sample size was chosen makes sense, it's probably worth trusting for evaluating that article. If it doesn't make sense, or if it is not explained

at all, it may indicate a problem for interpreting the study's results.

SPORTS MASSAGE

So what's going on with the research literature in sports massage? At a time when more and more athletes and trainers are using—and reporting benefits from—sports massage, why do so many studies report “no effect” of massage on recovery, like the Jönköping article mentioned above? What does that mean for sports massage?

Weerapong's (“The Mechanisms of Massage and Effects on Performance, Muscle Recovery, and Injury Prevention”) team looked at the evidence for massage in several areas of interest to athletes, including physiological and psychological indicators, performance improvement, injury prevention, and recovery. They found conflicting trends—many of the claims made for massage were not borne out by the results of studies, yet those studies also tended to have methodological problems.⁵ Additionally, they found that many of the claims made for neurological effects of sports massage, such as increasing neuromuscular excitability, have not been studied, so there is no evidence one way or the other for such claims. They did find evidence for other effects, including mechanical effects on muscles and psychological effects promoting recovery, and they call for more specific research, clarifying what massage techniques and outcome measures are examined, and avoiding the methodological flaws of previous studies. They call for research on the following specific questions:

- Can massage increase muscle blood flow, muscle temperature, neuromuscular excitability, or muscle flexibility?
- Can massage increase performance, such as sprinting, jumping, or endurance athletic events?

- What type of massage can produce benefits? How long should massage be applied? When should athletes receive massage?
- Are the effects of massage universal or are they specific to each massage therapist?
- Is the cost and time for massage appropriate when a warm-up or cool-down may be as, or more, effective?

The fourth research question raised by Weerapong's team is particularly intriguing, as it ties into research Moraska carried out on the effect therapist education had on outcomes after a race. He found that therapists with 950 hours of education achieved greater reduction in soreness than did therapists with fewer hours of education.⁶ This suggests that there may indeed be therapist-specific effects at some level that are escaping the studies, a consideration which gets back to the methodological trade-off mentioned above about the difference between studying the effects of a standard protocol versus studying what therapists really do (as difficult to replicate as that is).

Not only does this open the door to future research to explore what those therapist-specific effects may be, it also raises intriguing methodological questions about how to carry out those explorations. What trade-offs may be involved to preserve those effects while studying them, without having to sacrifice scientific rigor? It is a variation of the trade-off mentioned previously, about balancing the need for replicability in later studies with fidelity to the actual process of massage. It will remain an ongoing dialogue for some time to come.

The current state of the sports massage research literature is confusing and difficult to navigate. One of the reasons for this difficulty is the lack of methodological clarity, both within individual studies, and among them for comparison to each other.

This situation highlights the vital importance of solid methodology in research and the problems that can result when that methodology is not as strong as it could be. In order to gain a firmer shared understanding of what sports massage can offer, a crucial step is to reinforce a stronger methodology by clarifying what athletic outcomes are being looked for, what massage techniques are being examined, what therapist-specific factors need to be examined, and what study design factors—including sufficient sample size—are needed to ensure that the results are as valid as possible. **m&b**

6 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 researchingmassage@gmail.com with questions and comments.

NOTES

1. A. Moraska, “Sports Massage. A Comprehensive Review,” *Journal of Sports Medicine and Physical Fitness* 45, no. 3 (September 2005): 370–80.
2. L. Brosseau et al., “Deep Transverse Friction Massage for Treating Tendinitis,” *Cochrane Database of Systematic Reviews* 4 (2002): CD003528.
3. S. Jönköping et al., “Sports Massage After Eccentric Exercise,” *American Journal of Sports Medicine* 32, no. 6 (September 2004): 1499–503.
4. J. Hattan, L. King, and P. Griffiths, “The Impact of Foot Massage and Guided Relaxation Following Cardiac Surgery: A Randomized Controlled Trial,” *Journal of Advanced Nursing* 37, no. 2 (January 2002): 199–207.
5. P. Weerapong, P.A. Hume, and G.S. Kolt, “The Mechanisms of Massage and Effects on Performance, Muscle Recovery and Injury Prevention,” *Sports Medicine* 35, no. 3 (2005): 235–56.
6. A. Moraska, “Therapist Education Impacts the Massage Effect on Post-race Muscle Recovery,” *Medicine & Science in Sports & Exercise* 39, no. 1 (January 2007): 34–7.