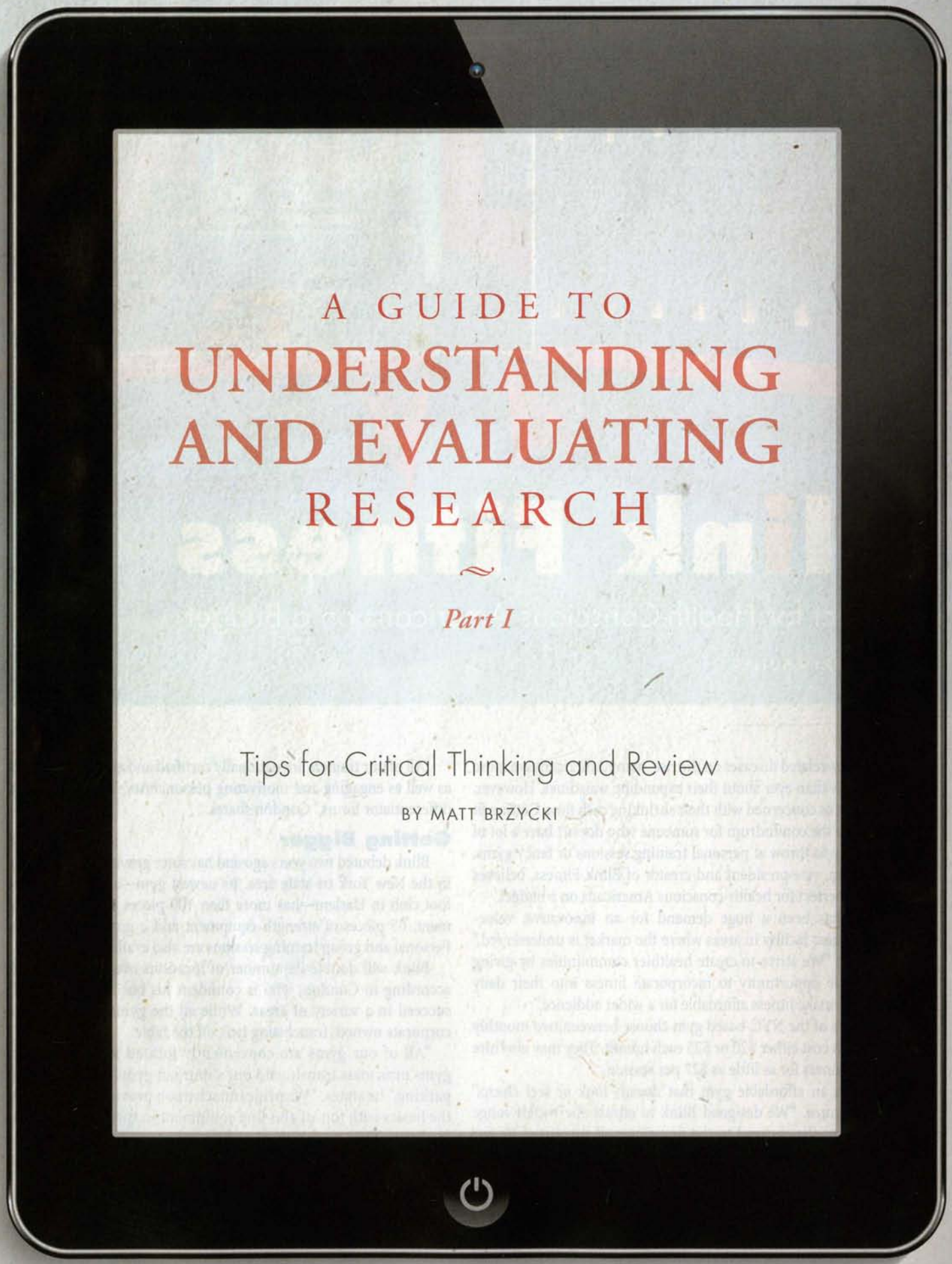

american FITNESS®
September/October 2014 $4.50
Teaching Pilates With POP
Cassey Ho
KICKBOXING in quadrants
Earn CEUs: Pilates-Based Postural Assessment
Healthy ALTERNATIVES to Sugar and ARTIFICIAL SWEETENERS



A GUIDE TO UNDERSTANDING AND EVALUATING RESEARCH

Part I

Tips for Critical Thinking and Review

BY MATT BRZYCKI

How many times have you heard on TV or read in the newspaper that "a recent study found..." or "new research shows..."? Do you blindly accept what's reported by the media as the gospel truth? If so, you shouldn't.

Case in point: A front-page article in *The New York Times* was titled "Study Cautions Runners to Limit Intake of Water."[12] Much of the article was based on a study that looked at 488 runners who provided blood samples and completed a questionnaire after finishing the Boston Marathon.[3] The researchers found that 62 of the 488 runners had hyponatremia, a condition that's characterized by a low concentration of sodium in the blood.

The primary risk factor for hyponatremia is thought to be an excessive intake of fluids (which is why hyponatremia is sometimes referred to as water intoxication). This dilutes the level of sodium in the blood, creating an electrolyte imbalance that impairs neural and muscular function. Most importantly, hyponatremia can be life-threatening.

The take-home message of the newspaper article was that people should limit their consumption of water when exercising. However, many individuals were likely frightened into the extreme, thinking that they should refrain from drinking fluids altogether.

According to the study, 168 of the 488 runners—more than one-third of them—drank so much fluid that they actually *gained weight during the marathon.* One individual was *9 pounds heavier* at the end of the race. Do you know how much fluid someone would have to drink to gain even 1 pound after running for several hours, let alone 9 pounds? The short answer is "a lot." To gain 9 pounds, an individual would have to drink about a gallon of fluid or more and that's in addition to replacing the weight that was lost from running for a few hours. What put those runners at risk for hyponatremia wasn't drinking fluids; it was drinking *excessive amounts of fluids.*

To make a long story short, it's a good idea for you to be skeptical whenever you hear or read a media report of a study that's related to health and fitness. In this two-part article, we will explore the best guidelines for evaluating research and point out important questions you should be asking prior to accepting research as legitimate.

TERMINOLOGY

Before discussing guidelines for evaluating research, it's important to understand some terminology that you'll likely encounter on a regular basis. Here are a few terms with which you should be familiar:

An **abstract** is a brief summary. It appears at the beginning of a study and may be about 200 to 250 words or less.

The **experimental group** consists of the subjects who receive a treatment or an intervention. Some studies have two or more experimental groups that receive different treatments.

The **control group** consists of the subjects who receive no treatment or intervention. This group is used as a point of comparison to see if any change by the experimental group was due to the treatment.

A **placebo** is a substance that contains no active ingredients. Often, it's a sugar pill and will be similar in appearance, taste and smell to the product that's being studied so that the subjects can't distinguish between the product and the placebo. Besides a pill, a placebo can also be a device or an item. For example, a study could compare the effects of a performance-enhancing bracelet to a fake (or "sham") that looks the same as the experimental bracelet.

A **variable** is something that can be changed. There are two types of variables: independent and dependent. Independent variables are the *presumed cause* and are *controlled* by the researchers. In a study of strength training, independent variables can be sets, repetitions, equipment or frequency of training. Dependent variables are the *presumed effect* and are *measured* by the researchers. In a study of strength training, dependent variables might be muscular size, muscular strength and percentage of body fat. To isolate the true cause of one or more effects, the researchers should manipulate only one independent variable and keep the

"THERE ARE MANY DIFFERENT TYPES OF STUDIES. BEING ABLE TO *DISTINGUISH* BETWEEN THEM WILL HELP YOU TO *UNDERSTAND* RESEARCH AND *EVALUATE* IT."

others the same. So a study that examines the effects of repetitions on muscular size and strength should assign different repetition schemes to the experimental groups without changing the number of sets, the type of equipment or any other independent variable.

Correlation refers to the association between two variables. A high correlation means that there's a strong association. However, correlation doesn't imply causation.

The term **significant** appears regularly in studies (or a derivative of the word such as *significantly*). In normal dialogue, significant means important; in statistical dialogue, it means probably true. It's used to describe the *amount of change* that's made by a group as well as the difference between two or more groups. When the amount of change is said to be significant, it implies that it's probably true that the amount of change was the result of the treatment rather than pure chance. When *the difference between groups* is said to be significant, it means that it's probably true that the difference was the result of the treatment rather than pure chance.

A **p-value** is the probability of obtaining a result and is the metric that researchers employ to determine whether the amount of change that's made by a group, or the difference between groups, is considered to be significant. Studies typically use a probability of 5%, which is a p-value of 0.05. When $p < 0.05$, it means that there's less than a 5% probability (5 of 100) that the amount of change or the difference was pure chance. Or, stated otherwise, it means that there's more than a 95% probability (95 of 100) that the amount of change or the difference was the result of the treatment.

Bias is any action that distorts the true findings. It's not unusual for a study to contain some degree of bias. The bias could be intentional or unintentional and can occur before, during or after the study. It manifests itself in many ways, including how the study is designed, how the data are interpreted and how the results are reported. Ironically, bias can also occur if a study *isn't* published. When studies that report negative or inconclusive findings don't get published, it's known as publication bias.

TYPES OF STUDIES

There are many different types of studies. Being able to distinguish between them will help you to understand research and evaluate it.

A **case study** is an in-depth review of one individual. This type of study must be interpreted with caution since the effects that are experienced by a sample of one (aka an N of one) can't be extrapolated to the general population.

In an **experimental study** (aka a treatment or an intervention study) the subjects are assigned to either an experimental group or a control group. As noted earlier, the experimental group receives a treatment and the control group does not.

In a **crossover study,** the subjects receive the experimental *and* control treatments. One group gets Treatment A followed by Treatment B while the other group gets Treatment B followed by Treatment A. The groups receive each of the assigned treatments for a designated period of time (which could be days or weeks, depending on what's being studied). In between the two treatments is a washout period that allows ample time for the first treatment to exit—or "wash out" of—the system so that there are no residual effects that might influence the second treatment. The subjects serve as their own controls in a crossover study.

A **survey** is a type of study in which quantifiable information is collected from a population or a subset of a population. Examples of surveys include questionnaires and interviews. Perhaps the most widely known is the U.S. Census.

A **cross-sectional study** is an observation of a population or a subset of a population at one specific point or "cross-section" of time. An example is measuring the body composition of a football team at the start of preseason camp (or a subset of the team such as the offensive linemen).

A **longitudinal study** is an observation of a population or a subset of a population at several different points over a long time. An example is measuring the body composition of a football team a few times over the course of a season—perhaps at the start of preseason camp, in the middle of the season and after the final game. A well-known longitudinal study is the Framingham Heart Study that has been ongoing since 1948.

A **position stand** (aka a position statement and consensus paper) is done by a panel of experts on behalf of an organization or a group. This yields a set of guidelines or "best practices" based on the available research that has been published on a specific topic. For instance, the Academy of Nutrition and Dietetics (formerly the American Dietetic Association) has a position stand on "Use of Nutritive and Nonnutritive Sweeteners"; and the American College of Sports Medicine has a position stand on "Exertional Heat Illness during Training and Competition."

In a **literature review** (aka a systematic review), researchers conduct an exhaustive analysis on all studies of acceptable quality that examined a specific topic. A review should include *all* relevant studies, not just the ones consistent with the perspectives of the researchers preparing the report.

A **meta-analysis** is a type of study that has been growing in popularity. Here, the results regarding studies that investigated the same thing are pooled together into one large study. So, for example, a meta-analysis that looks at 10 different studies with 50 subjects in each now has a compilation of data on 500 subjects. The criteria that are used to select studies for a meta-analysis shouldn't exclude any research that contradicts the perspectives of those researchers who are performing the meta-analysis.

QUESTIONS TO ASK

The mere mention of a study makes it sound as if there's credible evidence that a product or a program is effective. However the study may be poorly designed, irrelevant or taken out of context. And that's why it's imperative for you to "study the study" whenever possible so that you can separate *science fact* from *science fiction.*

But once you get your hands on a study—the *actual* study, not just an abstract or a media report—how do you evaluate it? Here are some questions that you should ask to determine if a study has any merit:

1. **Was the study published in a scientific journal?**

   Some studies haven't been published anywhere; others appear in nonscholarly publications. In either case, it means the study didn't go through a rigorous peer-review process in which experts in a related field (aka referees) do an impartial assessment of the manuscript to determine whether or not it should be published.

   With very few exceptions, magazines in bookstores and on newsstands are nonscholarly publications. Any claims about the safety or effectiveness of products and programs in these types of publications are largely based on anecdotal evidence, meaning that support is rooted entirely in personal experience, not scientific research. This isn't much better than someone in the gym who says that after using a specific product or doing a certain program, his arm circumference increased by 1.5 inches, his sprint time in the 40-yard dash decreased by 0.2 seconds and his bench press improved by 50 pounds. Basically, his "success story" is anecdotal evidence. The individual may have gotten bigger, faster and stronger, but there's no concrete proof that the changes were caused by the product or the program.

   Also be wary about information that's available on the Internet where quackery abounds. Although the emergence of the Internet has given people access to an unbelievable amount of information that's literally at their fingertips, not all of it is credible. Remember, any crackpot with a keyboard can—and does—post information on the Internet.

2. **Was the study properly designed?**

   Just because a study is published in a peer-reviewed journal doesn't guarantee that it's well designed. Consider this: A journalist submitted the same basic study of a "wonder drug" to 304 open-access journals.[5] The study carried the names of fictitious authors and affiliations and was intentionally loaded with design flaws that should have stuck out like a hippie at a Marine Corps Birthday Ball. Yet, more than half the journals accepted the spoof study for publication. (When accepted, the study was withdrawn so that it was never actually published.) Only 36 of the 304 journals gave review comments that the study contained design flaws. However, 16 of the 36 still accepted the study for publication.

   The gold standard in clinical research is a randomized, double-blind, placebo-controlled study. What does this mean?

"WHEN A STUDY IS TOO LONG, THERE'S ALSO A GREATER *POTENTIAL* THAT SOME SUBJECTS WILL NOT ADHERE TO THE *EXPERIMENTAL* PROTOCOL OR *WITHDRAW...*"

In a randomized study, the subjects are randomly assigned to groups—rather than selected or chosen for certain groups—in such a way that the physical/physiological profile and size of each group are roughly the same.

In a double-blind study, the researchers who are distributing the treatment and the subjects who are receiving the treatment are unaware—or "blinded"—as to who is getting what.

And in a placebo-controlled study, one group of subjects receives a treatment and another group—a control group—receives a placebo.

3. Who were the subjects in the study?

It's important to consider the subjects (or participants) who were studied. For a study to be relevant, the subjects should be somewhat similar to the population of interest.

Fucoxanthin, a compound that's found in edible brown seaweed, has been promoted as a substance for losing weight/fat. In one study, after just 4 weeks, two groups of female subjects who were fed diets that included lipid extracts of brown seaweed significantly reduced their body weight by as much as 36% more than those who were fed a control diet.[15] But—and this is a really big but—the subjects in the study were mice.

Responses that are experienced by animals can't always be generalized to humans. A classic example is resveratrol, a chemical found in red grapes and wine that has been touted as a substance for improving health and increasing longevity. In one study, mice that were fed a high-calorie diet plus resveratrol experienced significant improvements in life span and several markers of health more than two groups of mice that were fed either a high-calorie diet or a control diet.[4] But to get the same relative amount of resveratrol as the mice, two scientists estimated that a human would have to consume *about 333 glasses of red wine each day.*[11]

Even if a study does involve humans, they may only be representative of a very small segment of the population. Years ago, boron was promoted as a substance for increasing muscular size and strength less than three months after the publication of a study by the U.S. Department of Agriculture that showed subjects who received boron increased their testosterone by as much as 268%.[18] What wasn't emphasized about the study, however, was the fact that the subjects were 12 postmenopausal women, aged 48 to 82, whose testosterone was naturally low. Also, prior to taking the supplement, the women had been deprived of an adequate intake of boron for 119 days. Because postmenopausal women have a physical/physiological profile that's considerably different from the general population—or a subset of the population

such as competitive athletes—it's not reasonable to expect that others will experience the same effects from boron.

4. How many subjects were in the study?

The number of subjects in a study is known as the sample size. Determining what constitutes an adequate number of subjects requires the use of calculations or tables but, even then, is a matter of some debate. To a degree, it depends on the type of study and what's being investigated. A survey may require thousands of subjects, an experimental study may require 100 subjects and a crossover study even less. Commonly, a larger number of subjects will yield results that are more accurate and applicable. However, research with too many subjects might not be practical, placing a strain on time, budget and personnel.

The point, though, is that a study with a very small sample size—the so-called "small N"—tells us little about the safety and effectiveness of a product or a program. To be clear, a study that has a small sample size doesn't necessarily mean that it's flawed or should be disregarded. But it might be a big stretch to draw a valid conclusion from a small N.

5. Was the study of sufficient duration?

There's no clear guideline as to what constitutes an adequate length of a study; like sample size, it depends on the type of study and what's being researched. And like sample size, a study that's too long might not be practical, placing a strain on resources.

When a study is too long, there's also a greater potential that some subjects will not adhere to the experimental protocol or withdraw (or be "lost to follow-up"), which results in missing data. What's certain, though, is that long-term studies are needed to assess the safety and effectiveness of a product or a program.

In Part II of this article (to be featured in the November 2014 issue), we will discuss more questions you should ask when choosing reliable research, such as: Did the study find significant changes? Did the study show true cause and effect? And, did the data support the conclusions? See you then. **AF**

MATT BRZYCKI *is the Assistant Director of Campus Recreation, Fitness at Princeton University. He has authored, co-authored and edited 17 books including* A Practical Approach to Strength Training *(now in its fourth edition).*

REFERENCES:

1. AKL, E.A., ET AL. "POTENTIAL IMPACT ON ESTIMATED TREATMENT EFFECTS OF INFORMATION LOST TO FOLLOW-UP IN RANDOMISED CONTROLLED TRIALS (LOST-IT): SYSTEMATIC REVIEW." *THE BRITISH MEDICAL JOURNAL*, 344, NO. E2809 (MAY 18, 2012): 1-12.
2. AL-MARZOUKI, S., ET AL. "SELECTIVE REPORTING IN CLINICAL TRIALS: ANALYSIS OF TRIAL PROTOCOLS ACCEPTED BY *THE LANCET*." *THE LANCET*, 372, NO. 9634 (2008): 201.
3. ALMOND, C.S., ET AL. "HYPONATREMIA AMONG RUNNERS IN THE BOSTON MARATHON." *THE NEW ENGLAND JOURNAL OF MEDICINE*, 352, NO. 15 (2005): 1550-56.
4. BAUR, J.A., ET AL. "RESVERATROL IMPROVES HEALTH AND SURVIVAL OF MICE ON A HIGH-CALORIE DIET." *NATURE*, 444, NO. 7117 (2006): 337-42.
5. BOHANNON, J. "WHO'S AFRAID OF PEER REVIEW?" *SCIENCE*, 342, NO. 6154 (2013): 60-65.
6. BOUCHARD, C., ET AL. "FAMILIAL AGGREGATION OF VO_{2MAX} RESPONSE TO EXERCISE TRAINING: RESULTS FROM THE HERITAGE FAMILY STUDY." *JOURNAL OF APPLIED PHYSIOLOGY*, 87, NO. 3 (1999): 1003-08.
7. BRAY, G.A., NIELSEN, S.J. AND POPKIN, B.M. "CONSUMPTION OF HIGH-FRUCTOSE CORN SYRUP IN BEVERAGES MAY PLAY A ROLE IN THE EPIDEMIC OF OBESITY." *THE AMERICAN JOURNAL OF CLINICAL NUTRITION*, 79, NO. 4 (2004): 537-43.
8. FANELLI, D. "HOW MANY SCIENTISTS FABRICATE AND FALSIFY RESEARCH? A SYSTEMATIC REVIEW AND META-ANALYSIS OF SURVEY DATA." *PLOS ONE*, 4, NO. 5 (2009): E5738.
9. FRIEDMAN, L.S. AND RICHTER, E.D. "RELATIONSHIP BETWEEN CONFLICTS OF INTEREST AND RESEARCH RESULTS." *JOURNAL OF GENERAL INTERNAL MEDICINE*, 19, NO. 1 (JAN 2004): 51-56.
10. HUBAL, M.J., ET AL. "VARIABILITY IN MUSCLE SIZE AND STRENGTH GAIN AFTER UNILATERAL RESISTANCE TRAINING." *MEDICINE & SCIENCE IN SPORTS & EXERCISE*, 37, NO. 6 (JUN 2005): 964-72.
11. KAEBERLEIN, M. AND RABINOVITCH, P.S. "MEDICINE: GRAPES VERSUS GLUTTONY." *NATURE*, 444, NO. 7117 (NOV 2006): 280-81.
12. KOLATA, G. "STUDY CAUTIONS RUNNERS TO LIMIT THEIR WATER INTAKE." *THE NEW YORK TIMES* (APR 14, 2005): A1.
13. KRAEMER, W.J. "A SERIES OF STUDIES—THE PHYSIOLOGICAL BASIS FOR STRENGTH TRAINING IN AMERICAN FOOTBALL: FACT OVER PHILOSOPHY." *JOURNAL OF STRENGTH AND CONDITIONING RESEARCH*, 11, NO. 3 (1997): 131-42.
14. LESSER, L.I., ET AL. "RELATIONSHIP BETWEEN FUNDING SOURCE AND CONCLUSION AMONG NUTRITION-RELATED SCIENTIFIC ARTICLES." *PLOS MEDICINE*, 4, NO. 1 (JAN 2007): E5.
15. MAEDA, H., ET AL. "DIETARY COMBINATION OF FUCOXANTHIN AND FISH OIL ATTENUATES THE WEIGHT GAIN OF WHITE ADIPOSE TISSUE AND DECREASES BLOOD GLUCOSE IN OBESE/DIABETIC KK-AY MICE." *JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY*, 55, NO. 19 (SEP 2007): 7701-06.
16. MAURAGE, P., HEEREN, A. AND PESENTI, A. "DOES CHOCOLATE CONSUMPTION REALLY BOOST NOBEL AWARD CHANCES? THE PERIL OF OVER-INTERPRETING CORRELATIONS IN HEALTH STUDIES." *THE JOURNAL OF NUTRITION*, 143, NO. 6 (JUN 2013): 931-33.
17. MESSERLI, F.H. "CHOCOLATE CONSUMPTION, COGNITIVE FUNCTION, AND NOBEL LAUREATES." *THE NEW ENGLAND JOURNAL OF MEDICINE*, 367, NO. 16 (OCT 2012): 1562-64.
18. NIELSEN, F.H., ET AL. "EFFECT OF DIETARY BORON ON MINERAL, ESTROGEN, AND TESTOSTERONE METABOLISM IN POSTMENOPAUSAL WOMEN." *FASEB JOURNAL*, 1, NO. 5 (1987): 394-97.
19. PERLIS, R.H., ET AL. "INDUSTRY SPONSORSHIP AND FINANCIAL CONFLICT OF INTEREST IN THE REPORTING OF CLINICAL TRIALS IN PSYCHIATRY." *THE AMERICAN JOURNAL OF PSYCHIATRY*, 162, NO. 10 (2005): 1957-60.
20. SHAVE, R., ET AL. "THE EFFECTS OF SODIUM CITRATE INGESTION ON 3,000-METER TIME-TRIAL PERFORMANCE." *JOURNAL OF STRENGTH AND CONDITIONING RESEARCH*, 15, NO. 2 (2001): 230-34.
21. SMITH, M.M., ET AL. "CROSSFIT-BASED HIGH-INTENSITY POWER TRAINING IMPROVES MAXIMAL AEROBIC FITNESS AND BODY COMPOSITION." *JOURNAL OF STRENGTH AND CONDITIONING RESEARCH*, 27, NO. 11 (2013): 3159-72.
22. VAN ES, E.M. "HIGH-INTENSITY EXERCISE ON THE ROM TRAINING DEVICE IMPROVES VO_{2MAX} AND ENDURANCE CAPACITY IN UNTRAINED ADULTS." *MEDICINE & SCIENCE IN SPORTS & EXERCISE*, 39, NO. 5 (SUPPL.) (MAY 2007): S348-49.