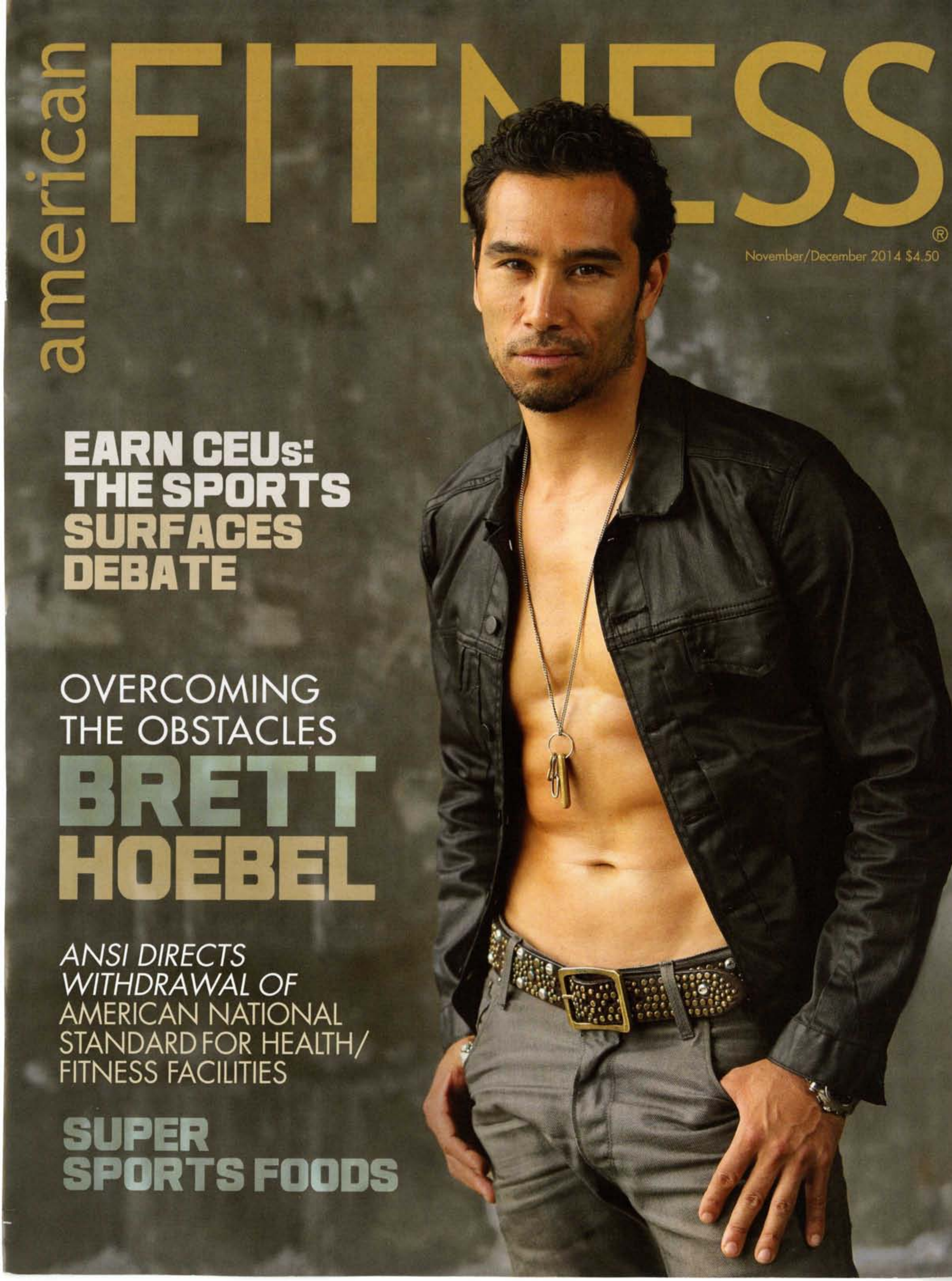
american
FITNESS®
November/December 2014 $4.50
EARN CEUs:
THE SPORTS
SURFACES
DEBATE
OVERCOMING
THE OBSTACLES
BRETT
HOEBEL
ANSI DIRECTS
WITHDRAWAL OF
AMERICAN NATIONAL
STANDARD FOR HEALTH/
FITNESS FACILITIES
SUPER
SPORTS FOODS

A GUIDE TO

UNDERSTANDING AND EVALUATING RESEARCH

Part II

Tips for Critical Thinking and Review

BY MATT BRZYCKI

Part one of this article in the May/June 2014 issue, discussed the best guidelines for evaluating research and answered important questions to consider before accepting research as legitimate. We continue now with question six.

6. Was the methodology (experimental protocol) unbiased?

A graphic example of biased methodology was illustrated in research utilizing 34 Division III football players.[13] In the study, the athletes were randomly assigned to one of two experimental groups: a single-set group or a multiple-set strength/power training group. In addition to being tasked with using different numbers of sets, the two groups were also assigned to utilize diverse repetition ranges, equipment, exercises, volume of exercises and different rest intervals between sets. A large number of independent variables makes it impossible to compare the results of two groups and draw valid conclusions. Indeed, there's no way to tell which variable was responsible for the effect.

Notably in this particular study, pre- and post-testing included the hang clean, an exercise that was featured in the program of the multiple-set group but not the single-set group. In effect, the multiple-set group practiced this exercise twice per week for 14 weeks (as well as a highly related movement—a mid-thigh pull—once per week) while the single-set group had no practice whatsoever. This gave the multiple-set group much greater familiarity with the hang clean and, as a result, placed the single-set group at a severe disadvantage when it came to being post-tested in that exercise.

7. Did the study account for low responders, high responders and other outliers?

An outlier is numerically distant from the rest of the data. The fact of the matter is that some individuals will show little or no improvement from a product or a program while others will show exceptional improvement.

Consider the HERITAGE (HEalth, RIsk factors, exercise Training And GEnetics) Family Study, in which 481 individuals from 98 two-generation families did aerobic training on stationary bicycles three times per week for 20 weeks.[6] On average, the subjects increased their oxygen intake by about 0.4 liters per minute (L/min). However, their response to training ranged from literally *no change* to more than 1.0 L/min, which was 2.5 times the average improvement.

Also worth noting is the FAMuSS (Functional Polymorphisms Associated With Human Muscle Size and Strength) Study, wherein 585 individuals did strength training with their nondominant arm for 12 weeks.[10] On average, the subjects increased their cross-sectional area by 3.2 centimeters squared (cm^2), their isometric strength by 7.5 kilograms (kg) and their dynamic strength by 3.9kg (all measures in their trained biceps). However, their response to training in those three variables ranged from a *decrease* of $0.5cm^2$ to an increase of $13.6cm^2$; a *decrease* of 15.9kg to an increase of 52.6kg and *no change* to an increase of 10.2kg, respectively.

The problem with outliers is that they can skew the data and produce false or misleading conclusions. For instance, suppose that a study collected 10 data points: Nine subjects had 10s and one subject had 60. The average of those 10 data points is 15 but 9 of the 10 subjects—90% of them—were well below average

"IT'S ALSO ***IMPORTANT*** TO UNDERSTAND THAT A CHANGE OR DIFFERENCE CAN BE ***'STATISTICALLY SIGNIFICANT'*** BUT NOT ***'CLINICALLY SIGNIFICANT.'*** "

with 10s. So the 60—an outlier—distorted the data. Outliers can really wreak havoc on the data in studies that have a small number of subjects.

8. Did the study find significant changes?

Imagine a study in which the subjects are randomly assigned to two groups: One group receives a supplement and the other group receives a placebo. Both the supplement and the placebo could produce a significant change in some dependent variable—such as muscular size or strength—without there being a considerable difference between the supplement and the placebo. Thus the group that received the supplement might experience a greater amount of change than the group that received the placebo, but the disparity between the two groups might not be large enough to conclude that the supplement is superior to the placebo; rather, the difference may be due to pure chance.

It's also important to understand that a change or difference can be "statistically significant" but not "clinically significant." In other words, the change or difference may be so small that it has no practical consequence in the real world. For example, what if a study found that after 26 weeks, the subjects who received a product lost significantly more weight—three times as much weight, in fact—as the subjects who received a placebo? At first blush, that sounds quite impressive but, what if "three times as much" meant that the product produced a weight loss of three pounds while the placebo produced a weight loss of one pound? A difference of two pounds between the product and the placebo no longer sounds impressive—especially after 26 weeks—and has little relevance to people trying to lose weight.

9. Did the study show a true cause and effect?

An oft-stated maxim in statistics is that "correlation doesn't imply causation." In other words, a correlation between two variables doesn't mean that one caused the other; it could simply be a coincidence. Nonetheless, this shouldn't be dismissed completely since there could very well be an association between the two variables.

A physician once made the case that there's a "powerful" correlation between the amount of chocolate that a country consumes per capita and the number of Nobel Prize winners that a country produces (which was used in the article as a proxy for cognitive function).[17] His brief article—which was published in a prestigious medical journal—had some drawbacks. For one thing, the data used the consumption of chocolate by individual countries, not individual Nobel laureates. Also, the data showed the intake of chocolate over a two-year period but looked at more than 100 years of Nobel laureates.

In response to the article, researchers in Belgium pointed out that there was a high correlation between the number of IKEA® furniture stores that a country has per capita and the number of Nobel Prize winners that a country produces.[16] And the correlation was actually *higher* than that of chocolate. So just because there's correlation doesn't mean there is causation.

Similarly, there has been an ongoing attempt to link high-fructose corn syrup (HFCS) with obesity. The roots of this movement can be traced back to researchers who noted that over a 30-year period, as the consumption of HFCS increased, so did the rate of obesity.[7] They concluded that HFCS causes obesity. But again, correlation doesn't imply causation. Over the past 30 years, as property taxes increased, the obesity rate did too. Does this mean there's a correlation between property taxes and obesity? There are far too many variables at play in the obesity epidemic to single out one variable as the archvillain.

So if correlation doesn't imply causation, what does? Well, the best way to show causality is through a randomized, double-blind, placebo-controlled study. In other words, it's through a study that involves an *intervention,* not an *observation.*

10. Did the study have dropouts and, if so, howwere the missing data handled?

Many studies—particularly those that involve a large number of subjects and are of long duration—will have dropouts. Subjects withdraw from studies for a variety of reasons. For instance, they might experience a major change in life (such as moving away from the area or losing a job), sustain an illness or incur an injury (which could be related or unrelated to the study). Or some subjects may simply choose to no longer participate. Regardless of the reason, when subjects are "lost to follow-up," it results in missing data. And missing data can be related to the safety and effectiveness of a product or a program.

Of no small importance would be dropouts that are a direct result of a treatment, something that could bias the outcome of a study and distort the conclusions. For example, if a large number of subjects withdrew from a study because they experienced an adverse effect, then this could have a major impact on the outcome and its interpretation, especially if the dropouts weren't evenly distributed among the experimental and control groups.

In one literature review, researchers looked at 235 studies that found significant results.[1] Dropouts were reported in 191 of the 235 studies. In most cases, the studies—which were published in five prestigious medical journals over a three-year period—had no explanation as to why the subjects were lost to follow-up. So the literature review used plausible assumptions and found that as many as 33% of the studies would no longer have significant results when the dropouts were considered. It's pretty clear, then, that dropouts can have an enormous influence on outcomes.

The researchers should note how they accounted for the missing data from the subjects. Yet in the aforementioned literature review, 20% of the studies didn't report how the missing data were handled.

There's no standard way to account for missing information. One approach, however, is the Last Observation Carried Forward (LOCF) method. As the name implies, when a subject withdraws from a study, the data that are last recorded are "carried forward" to the end of the study.

11. Did the data support the conclusions?

Researchers sometimes draw conclusions that aren't supported by the data. This could be intentional or unintentional and can occur for a number of reasons. One typical reason is overgeneralizing the results. An example would be a study that found doing sets of lower repetitions of the leg press was significantly better than doing sets of higher repetitions in the performance of a one-repetition maximum, then broadly determining that lower repetitions are superior to higher repetitions.

Other reasons the data fail to support the conclusions have to do with the design of the study. This includes utilizing data that are inadequate (i.e., as from having a small number of subjects) and using poor methodology (such as from having more than one independent variable).

Remember, the conclusions are where researchers have the opportunity to put their own spin on the data. As "spin doctors," they can choose to employ or ignore information, depending on whether it supports their position. Therefore, it's a good idea to review the design of the study with great care to see if the data really defend the conclusions.

12. Did the study report spectacular results?

As they say, "If it sounds too good to be true, it probably is."

According to the advertisements for one exercise machine, using it for "exactly four minutes per day" will produce the "combined results" of 20 to 45 minutes of aerobic training, 45 minutes of strength training and 15 to 20 minutes of stretching. In other words, four minutes on the machine is supposedly the equivalent of 80 to 110 minutes of physical activity.

One study of the machine was conducted in the Netherlands and carved into two poster presentations shown at an annual sports medical meeting. The presentations were published as two brief abstracts in a peer-reviewed journal.[22] In the study, 16 subjects used the machine three times a week for eight weeks. Each session involved eight minutes of training, four minutes with the upper body followed by four minutes with the lower body. Over the course of eight weeks, each subject trained for a total of three hours and 12 minutes. In a cycling test to exhaustion, the subjects increased their endurance by 72%, from 14:51 to 25:31

and no, that's not a typo. This improvement is so spectacular that it must be interpreted with extreme caution.

13. Was there a selective reporting of the results?

There's an old saying that if you torture the data long enough, you can make it confess to anything. Unfortunately, the selective reporting of results—a form of bias—isn't unusual. Indeed, a study that looks at enough dependent variables is bound to find one positive outcome.

In one literature review, researchers examined selective reporting in 37 studies that were published in a prestigious medical journal.[2] The review compared the summary protocols of the studies when they were first accepted for publication to the studies after publication. In 11 of the 37 studies, the researchers found "major differences" in the primary outcome between the accepted and the published versions. Eight studies introduced a new primary outcome. No reasons were given for the changes.

On a related note, the same outcome can result in two conclusions that are quite similar but with vastly different messages. For instance, suppose that there was a study in which one group did slow-speed repetitions and another did fast-speed repetitions. After 12 weeks of training, both groups significantly increased their vertical jump and there was no significant difference between the groups. Based on these outcomes, a researcher who personally favors fast-speed training might determine that doing slow-speed repetitions was no better than doing fast-speed repetitions for improving the vertical jump; a researcher who personally favors slow-speed training might conclude the opposite. Technically, both conclusions are correct and based on the same outcome yet convey different meanings.

Researchers should report *all* results, not just those in which their findings match their feelings. A selective reporting of results is unscientific and, frankly, unethical.

14. Did the study find any adverse effects or reportany injuries?

A study might show that a product or a program lives up to its hype but it is important to know whether or not it resulted in any adverse effects or injuries.

Sodium citrate is a supplement that has been promoted for improving endurance. In a crossover study, nine elite athletes ran 3,000 meters on two separate occasions, one after receiving sodium citrate and another after receiving sodium chloride (table salt).[20] The subjects ran 3,000 meters significantly faster—by an average of about 10 seconds—after consuming sodium citrate. And don't forget, these were elite athletes so a 10-second improvement in a 3,000-meter run is pretty substantial. That's the good news. The bad news is that when using sodium citrate, eight of the nine athletes experienced gastrointestinal distress.

It's equally important to consider injuries. Researchers at Ohio State University had 54 subjects perform a program that was based on a popular fitness regimen.[21] After 10 weeks, the subjects had significant improvements in their oxygen intake and percentage of body fat. But nine of the 54 subjects dropped out of the study due to overuse or injury. That's one in six subjects. Those aren't great odds. To their credit, the researchers expressed concern about the injuries in the discussion section of the study.

Even if the study didn't find any adverse effects or report any injuries, the duration might have been too short. Long-term studies are needed to assess the safety and effectiveness of a product or a program. Remember, individuals might use a product for months or years, not days or weeks. So it's all well and good if a product doesn't produce any adverse effects after four weeks. But after six months, will you grow a third eye in the middle of your forehead or sprout hair on your knuckles?

Also take into account that some studies don't investigate negative effects or injuries. In other cases, such problems go unreported.

Along these lines, individuals often take more than one product at a time. Combining two or more products can yield adverse effects.

15. Was the study funded and, if so, by whom?

Nowadays, most journals require researchers to disclose all sources of funding and professional relationships with any company or organization that may benefit from favorable outcomes…and for good reason.

Studies on a product are often funded by manufacturers of the same product. When a manufacturer pays to have its own product investigated, it increases the possibility that the study could be biased in some way. Needless to say, research funded by manufacturers with a direct financial interest in the outcome should be viewed with suspicion.

Industry financing has been investigated extensively and the results are unequivocal: The outcome of a study tends to favor the funder. An examination of 206 studies found that research with industry funding was about four to eight times more likely "to be favorable to the financial interests of the sponsoring company" than that without industry funding.[14] An analysis of 162 reports that were published in four psychiatric journals found that studies with a conflict of interest were nearly five times more likely to have a positive outcome than studies without a conflict of interest.[19] That's nothing: An analysis of

"AN ANALYSIS OF ***162 REPORTS*** THAT WERE PUBLISHED IN FOUR PSYCHIATRIC JOURNALS FOUND THAT STUDIES WITH A CONFLICT OF INTEREST WERE NEARLY ***FIVE TIMES MORE LIKELY*** TO HAVE A ***POSITIVE*** OUTCOME…"

398 studies published in two medical journals found that researchers who had a conflict of interest were 10 to 20 times more likely to present a positive outcome than those without one.[9] Clearly, the funding can affect the finding.

Try to determine if one or more researchers have any financial ties to the sponsor of the study. This includes being a paid employee or consultant, serving on an advisory board, receiving honoraria, owning stocks and having a patent agreement. If there's a conflict of interest, you must decide whether it could have influenced (biased) the outcome of the study. Note: Just because there's a conflict of interest doesn't mean that a study is biased.

16. Were the results of the study replicated by other researchers in other laboratories?

There are instances where several studies show the same results but are conducted by the same researchers at the same laboratories. For there to be compelling evidence, similar results need to be found by different researchers at different laboratories. When different groups of researchers investigate the same topic and show comparable results, it also adds to the body of existing evidence.

Furthermore, similar results produced by different researchers lessen the chance that the studies were biased toward a specific outcome. Otherwise, we are left to wonder, "Did the doctor 'doctor' the data?" In fact, cases of fraud are often uncovered when other laboratories are unable to replicate the results of a study.

Fraud is more prevalent in research than most people think. According to a meta-analysis that included 18 studies, nearly 2% of researchers admitted to having "fabricated, falsified or modified data or results at least once."[8] And a little more than 14% of researchers conceded that they had "personal knowledge of a colleague who fabricated or falsified research data or who altered or modified research data." With the volume of studies that is published, those numbers are quite disturbing.

THE BOTTOM LINE

Fitness professionals must stay abreast of the latest research in health and fitness. To determine the true value of a study, its content must be fully vetted. Having a basic understanding of terminology and the types of studies will improve your ability in evaluating research. **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).*

Note: *For a complete listing of references used for this article, please refer to Part I published in the September 2014 issue.*