HARD TRAINING

...To Provide Information and To Stimulate Thought On The Art of Strength Training
"Read Not To Believe... But To Weigh and Consider... —A.S.

No. 12 2000

REALISTIC RESULTS?

By Dr. Ken E. Leistner

One of the most provocative chapters in the classic Arthur Jones publication, Nautilus Training Bulletin Number One, reveals Arthur's conclusions and observations regarding the eventual results of proper strength training. "Average Expectations From Training" rankled some, stimulated others. Remember that Arthur's interpretation of "proper" differed from most when this "must have" volume was self-published in 1970. To be very truthful, Arthur's interpretation of "proper" training differed from almost everyone else upon the publication of Bulletin Number Two in 1971, and to this very day, I often chuckled when I thought of Arthur banging away at his typewriter, putting what often appeared to be a seething, barely harnessed energy to work, trying to somehow put into the written word, his very distinct theory of proper and effective training. I imagined Kim Wood sitting next to him, reaching a hand out to accept each subsequent page, having proofread the one before. This, more or less, was how the Bulletin was completed. I think too of Arthur's conversations, sometimes long affairs we had in the small office that was located in the training area of the factory/showroom of the Lake Helen facility. In every case, Arthur would be very clear, almost physical in his descriptions of what proper training was.

Simply put, Arthur believed one had to train hard. His version of hard was not like that of others. Arthur trained me numerous times. I served as a guinea pig, demonstrator of the equipment, demonstrator of a "typical training session" when coaches came into town as potential customers or when I assembled equipment after making the truck borne deliveries, and sometimes it seemed, just to provide him with someone to whip up on! I had trained "hard" as a football player, powerlifter, and interested trainee and Arthur, I believe, didn't dislike me upon first introduction because I did in fact, train hard in his estimation. To this day, few understand the meaning of that expression. I have had more athletes, competitive bodybuilders and lifters, "general training types", and aspiring actors tell me that they "train harder than anyone else" and certainly "as hard as (I) will need them to" and it is infrequently seen that anyone truly does. In his hey day, when Arthur was focused on turning out great equipment and getting the best results possible, he demanded an awful lot of each trainee. Few realize how strong Kim Wood was (and still is); how strong Tom Laputka was (and still is); how truly strong Casey Viator was (and I'm sure still is); how strong Dennis and Walter Anderson were, unsung training partners and "the guys who pushed" Casey to his limits. If you trained "realistically hard" you had to improve, assuming you ate and rested enough to come back and do it another day.

Thus, Arthur's perspective was certainly skewed. He knew what hard training was, what is should look like when someone did it, and what that would lead to with consistency over time. The Bulletin chapter was the result of those specific observations. Over the past decades, I have come to realize that few will train as hard as the core group of Nautilus did in the early 1970s. Few will train as hard as is necessary to produce optimal results, "all of the time" or even "most of the time". Some will do so but rarely. This is not an indictment of the training community but rather, what I think is an accurate observation. This will obviously affect typical training results and alter what is actually realistic. Many authors in this field have been clear than not everyone can become a Mr. America type physique or a world champion lifter or strongman competitor. This seemingly innocuous statement has taken on a life of its own, with one extreme defending the statement and noting that as "hard gainers" or as one who has minimal genetic advantages, they are doomed to a life of undersized and under strengthened existence despite their best efforts. The other extreme is just as adamant that by rolling up one's psychological sleeves, getting in there, being as "tough and manly" as possible, any perceived genetic limitation can be overcome. Both utilize examples of under development or over development on some well known figure in the field to support their stance. Obviously,

joyable and impressive ways.

Hammer Strength, phone toll free 1-877-543-1123, athletic sales office at 2245 Gilbert Ave., Cincinnati, OH 45206 and ask for my favorite ladies, Gail or Michelle. Computer site at hammerstrength. com for complete line of plateloading and the new selectorized Hammer Strength equipment.

Richard and Bert Sorin, phone toll free 1-877-767-4639, Box 121, Irmo, SC 29063. Computer site at sorniex.com for a supermarket of forearm and other strength training items.

Iron Mind Enterprises, phone 1-530-265-6725, Box 1228, Nevada City, CA 95959. Computer site at ironmind.com for great forearm and grip items.

PDA, John Syzmanski for unusual grip and farmer's walk devices, phone 1-864-963-5640, 104 Bangor St., Mauldin, SC 29662, Send e-mail inquiries to pda@fractionalplates.com

---

FLAWS IN RESEARCH DESIGN AND INTERPRETATION - PART I

Matt Brzycki
Coordinator of Health Fitness, Strength and Conditioning
Princeton University

The strength and fitness community has been bombarded with results from a growing number of research studies that supposedly prove one point or another. Before the results of studies can be accepted and applied, it must be determined that they were conducted in a scientifically legitimate manner. Quite often, many studies are infected with numerous design flaws thereby rendering their results to be questionable at best and meaningless at worst. Indeed, some studies have design flaws that would not be worthy of a passing grade in a high school science project yet they are somehow capable of passing the supposed peer-review process of a scientific journal. Other studies have been done in a fashion that is scientifically acceptable but their results have been taken out of context or broadly generalized.

A multitude of studies in strength training has manipulated different training variables in an effort to determine an optimal program or method. One of the most popular ways has been to compare groups using different numbers of sets (and repetitions) of an exercise or exercises. For years, it was generally accepted that performing multiple sets of an exercise was better than single sets. Carpinelli and Otto (1998) destroyed this belief with an astonishing discovery: Their comprehensive literature review of all relevant research that examined different numbers of sets — 35 studies which dated back to 1956 — showed that there were no significant differences between single- and multiple-set training in all but two studies: Berger (1962) and Kramer and others (1997). In a later review, Carpinelli (1999) noted two additional studies that found no significant differences between single- and multiple-set training.

THE TERM "SIGNIFICANT"

During discussion of the studies, a term that appears frequently is "significant" (or a derivative of the term such as "significantly"). In normal dialogue, "significant" means "important"; in statistical dialogue, "significant" means "probably true." The term "significant" is used to describe the magnitude (or size) of a change as well as the difference between two or more groups. When the magnitude of change is said to be "significant" it means that it is "probably true" that the magnitude of change was the result of the training protocols rather than pure chance; when the difference between two or more groups is said to be "significant," it means that it is "probably true" that the difference was the result of the training protocols rather than pure chance.

The degree to which a relationship is "probably true" is based upon the "statistical significance" or "p-value." If a relationship is said to be "significant" at the $p<0.05$ level, it indicates that there is more than a 95% probability (or chance) that the result was due to the training protocol; if a relationship is said to be "significant" at the $p<0.01$ level, it indicates that there is more than a 99% probability (or chance) that the result was due to the training protocol. Results with a high probability of occurring as a result of the training protocol are said to be "statistically significant." A p-value of 0.05 is usually considered to be borderline statistically significant.

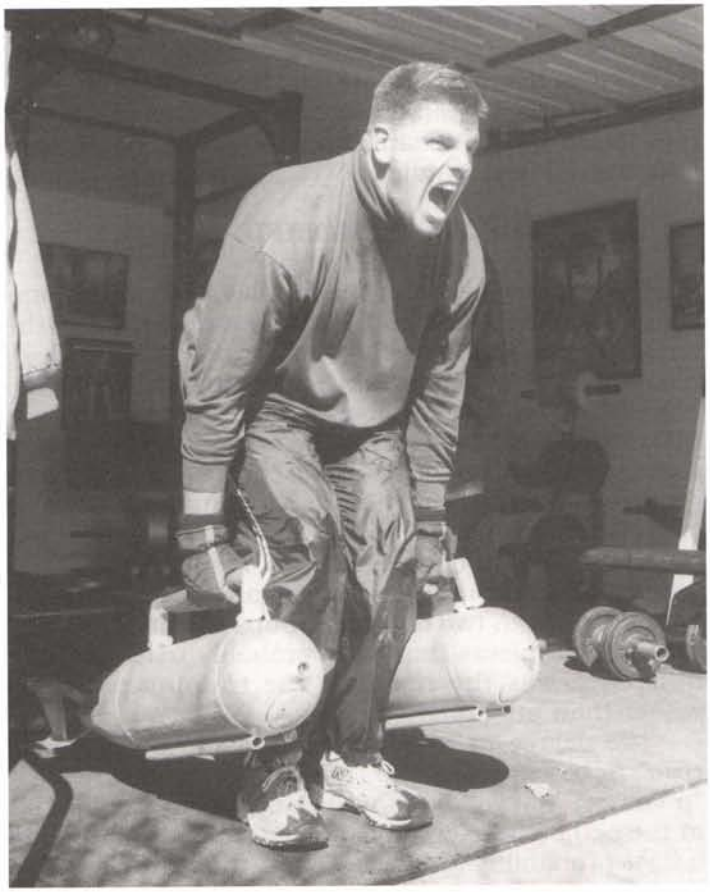
THE BERGER AND KRAMER STUDIES

What follows is a closer examination of the studies by Berger (1962) and Kramer and his co-workers (1997).

The "Optimum Combination of Sets and Repetitions"?

In this 12-week study, Berger (1962) suggested that a greater volume of training — that is, "more sets, more repetitions per set and more total repetitions each training session" — produced a greater improvement in strength. His study has been cited countless times as proof that multiple sets are better than single sets and a higher volume of training is better than a lower volume of training. However, an in-depth look at his research reveals some interesting — and conflicting — information.

The study involved 177 males who were freshmen and sophomore students enrolled in nine weightlifting classes at the University of Illinois. In addition to their regular weightlifting program, the classes trained three times per week (on Monday, Wednesday and Friday)



Brian Saxton at 6'5" 260 lbs. tight end for the Atlanta Falcons performs a set of Bomb deadlifts

for 12 weeks with the bench press (using a barbell). In the study, the students were assigned into one of nine groups who used different combinations of one, two and three sets and two, six and ten repetitions. In other words, the nine groups used the following protocols (sets x repetitions): 1 x 2, 1 x 6, 1 x 10, 2 x 2, 2 x 6, 2 x 10, 3 x 2, 3 x 6 and 3 x 10.

The groups were tested in their one-repetition maximum (1-RM) bench press (using a barbell). Testing occurred five times during the study: at 0, 3, 6, 9 and 12 weeks. At the end of the study, the researcher concluded that the "optimum combination of sets and repetitions" for the development of strength was 3 x 6.

Some comments about the design and administration of this study:

- In a scientific study, it is important to use a sample size that is appropriate. On the one hand, the use of 177 subjects certainly sounds acceptable — especially considering that this study involved nine experimental groups; on the other hand, such a large number of subjects raises a question as to whether or not there was adequate supervision. Since only one author is named, it is assumed that this study involved only one researcher. In order to assure that the training was performed exactly as prescribed, close supervision of every repetition and set is required. Having 177 subjects do the bench press three times per week for 12 weeks amounts to 6,372 individual sessions — with roughly two thirds of those sessions involving 2 - 3 sets. Question: How can one researcher adequately supervise 177 subjects so as to make certain that they used their assigned set/rep protocol over the course of 12 weeks in the manner prescribed to assure the scientific purity of the study?
- In a scientific study, it is important to control as many known variables as possible. The fact that the test subjects were involved in another form of strength training outside the study — that is, a weightlifting class — presents a huge extraneous variable that was not controlled. This inconstant could have had an enormous impact upon the results of the research.
- In a scientific study, the characteristics of the groups — such as physical profiles and performances — should be balanced or "matched" as much as possible so that they are not distinguishable. In this study, however, the nine groups were not equated for strength before training began which could have tainted the results of the research.
- The researcher chose the bench press because its execution "did not require any particular skill." Anyone with any experience lifting weights knows that a certain amount of skill is required in bench pressing a barbell. Moreover, the 177 subjects undoubtedly had varying degrees of previous experience (and prior skill) with the bench press that

would have affected the results of the research.

- During the 1-RM testing, each subject "rested 2 - 3 minutes between attempts." There is a big difference between resting two minutes and three minutes. What if a subject rested two minutes between attempts in an earlier test and three minutes in a later test? Or just the opposite?

Some comments about the test results and conclusions:

- From weeks 7 to 9, the group who did 1 x 10 had the greatest rate of improvement of the nine groups (5.9%). During that same three-week period, the group who performed 1 x 10 increased their 1-RM bench press by an average of 8.2 pounds while the average of the other eight groups was 5.1 pounds. It would not be expected that the 1 x 10 group would have the best performance of the nine groups if, as the researcher suggested, a greater volume of training produces greater results.
- The researcher noted that "improvement rates were practically the same during the last 3 weeks of training." True, but a closer inspection of the data reveals that during the last three weeks of training — weeks 10 to 12 — the group who did 1 x 6 experienced the greatest rate of improvement of the nine groups (4.8%). The second-best rate of improvement was by the 3 x 6 group (4.0%).
- From weeks 10 to 12, the group who performed 1 x 6 increased their 1-RM bench press by an average of 7.2 pounds. The average of the other eight groups was 4.3 pounds. During that same three-week period, the group who did 1 x 10 had a greater rate of improvement (3.6%) than the group who did 3 x 10 (2.6%).
- All groups significantly increased their average 1-RM strength throughout the 12-week training period ($p<0.05$).
- According to the data, the group who did 3 x 6 had the greatest improvement of the nine groups. It should be noted, however, that the second-best performance was recorded by the group who did 1 x 6. It would not be expected that the 1 x 6 group would perform better than seven of the other eight groups if, as the researcher suggested, a greater volume of training produces greater results.
- Even though the improvement of the group who did 3 x 6 was greater than the group who did 1 x 6, it was not significantly greater. In effect, the 3 x 6 group performed three times as many sets (or 200% more) than the 1 x 6 group without obtaining a significantly greater increase in their 1-RM bench press.
- The average improvement in 1-RM strength every three weeks by the three-set groups (i.e., 3 x 10, 3 x 6 and 3 x 2) was 7.95 pounds. The next best improvement was by the one-set groups (7.00 pounds) followed by the two-set groups (6.85 pounds). It would not be expected that the one-set groups would perform better than the two-set groups if, as the researcher suggested, a greater volume of training produces greater results.
- The groups who did six repetitions (i.e., 1 x 6, 2 x 6 and 3 x 6) had a significantly greater improvement in 1-RM strength than the groups who did two repetitions but not significantly better than the groups who did ten repetitions. That is, there was no significant difference in the improvement in 1-RM strength between the groups who used six repetitions and the groups who used ten repetitions. This is actually quite interesting from the standpoint of training specificity since the testing involved obtaining a 1-RM. In this case, it would be expected that the groups who did two repetitions would perform better than the groups who did higher repetitions.
- The researcher noted that the "results indicated that more sets, more repetitions per set and more total reps each training session resulted in greater improvement in strength." Yet, the order of the groups in terms of improvement (from greatest to least) were as follows: 3 x 6, 1 x 6, 2 x 10, 3 x 2, 3 x 10, 2 x 6, 1 x 10, 1 x 2 and 2 x 2.
- This study involved one exercise: the bench press. Therefore, a better conclusion would have been that in this study, the "optimum combination of sets and repetitions" was 3 x 6 for the bench press but not significantly better than 1 x 6.
- The results of this study — that 3 x 6 was the "optimum combination of sets and repetitions" — were never replicated by the researcher or anyone else.

"Multiple Sets Increase the 1-RM Squat at a Faster Rate Than 1 Set to Failure"?

The second study noted in the literature review by Carpinelli and Otto (1998) to show the superiority of multiple sets was that of Kramer and others (1997). This 14-week study initially involved 53 males (average age 20) but 10 "withdrew before completion due to illness or injury unrelated to the study." The remaining 43 subjects were randomly assigned into one of three groups: a single-set group, a multiple-set group or a varied multiple-set group. The three experimental groups exercised three times per week for 14 weeks. All groups performed the same workouts which consisted of the following exercises: squat, push press, bench press and crunch on Mondays and Fridays; pull (from mid-thigh), leg curl, bent-over row and crunch on Wednesdays. However, the groups did different sets and repetitions. For each exercise, the single-set group did 1 x 10 with 50% of their target weights followed by 1 x 8 - 12-RM for all 14 weeks. The multiple-set group did 1 x 10 with 50% of their target weights and 1 x 10 with 75% of their target weights followed by 3 x 10 with their target weights for all 14 weeks. In the "major" exercises — which were not specified by the researchers except for the squat — the varied multiple-set group performed varying sets and repetitions at a target weight (1 x 10 in week 1, 3 x 5 in weeks 2 - 3, 3 x 3 in weeks 4 - 5, 1 x 10 in week 7, 3 x 5 in weeks 8 - 9, 3 x 3 in week 10, 1 x 10 in week 11, 3 x 5 in week 12, 3 x 3 in week 13 and 3 x 2 in week 14); in the "assistance" exercises — which were speci-

fied by the researchers as the crunch, leg curl and bent-over row — the varied multiple-set group performed 3 sets of varying repetitions at a target weight (3 x 10 in weeks 1 - 2 and 3 x 5 in weeks 3 - 14). Although there was no indication, it is assumed that the varied multiple-set group did 1 - 2 warm-up sets prior to their assigned protocols. The subjects chose their own rest periods between sets and exercises that was — according to the researchers — "typically 2 - 3 min." The subjects in the single-set group selected their own target weights; the "initial training loads" for the multiple-set groups were set by the researchers and "adjusted throughout the study." The multiple-set groups used "heavy loads" on Mondays and "light loads" on Fridays (10% less than that used on Mondays). The single-set group did each exercise to the point of fatigue; the multiple-set groups did not.

The groups were tested in their 1-RM squat, body composition and body mass. Testing occurred three times during the study: at 0, 5 and 14 weeks.

Some comments about the design and administration of this study:

- The experimental groups were not equated for strength before training began which could have tarnished the results of the research. The researchers stated that there were "no significant differences . . . between groups for initial 1-RM squat." Yet, the initial levels of 1-RM squat for the three groups varied by as much as 12.89% (from 98.5 kilograms to 111.2 kilograms).
- The researchers stated that the subjects were "randomly assigned to one of three experimental groups." A random assignment of subjects suggests that the size of each group is equal or nearly equal. Yet, the groups differed in size by as much as 23.08% with 16 in the single-set group, 14 in the multiple-set group and 13 in the varied multiple-set group. This unequal distribution of the subjects — which might be due to the fact that 10 of the original 53 subjects withdrew from the study — may have influenced the results of the research.
- The researchers stated that "subjects had to complete 90% of the training days to be included in the results of the study." In 13 weeks of actual training, this means that a subject could miss as many as 4 of the 39 total workouts and still be included in the results of the study. This uncontrolled variable — that is, the fact that the subjects performed anywhere from 35 - 39 workouts — could have affected the results of the research.
- Technically, the single-set group did two sets of each exercise (one being a "very light" warm-up with 50% of their target weights for 10 repetitions) and the multiple-set group did five sets of each exercise (two being with 50% and 75% of their target weights for 10 repetitions each).
- The researchers stated, "The subjects chose their own rest periods" which were "typically" 2 - 3 minutes between sets and exercises. Later, they stated that the time taken by the single-set group to complete their training sessions "ranged from 30 to 40 min and was typically 40 - 50 min for the multiple-set groups." The single-set group did 8 sets per workout which means that the average time between their sets was about 3.75 - 5 minutes (including the time taken to perform the exercise); the multiple-set group did 20 sets per workout which means that the average time between their sets and exercises was, amazingly, about 1.5 - 2 minutes (including the time taken to perform the exercise). There is no indication that the subjects used different speeds of movement to perform their repetitions. Therefore, it is assumed that the single-set and multiple-set groups took roughly the same amount of time to complete their sets. This uncontrolled variable — that is, the fact that the subjects in the single-set and multiple-set groups took much different amounts of rest between their efforts — could have affected the results of the research.
- The researchers stated that "All training sessions . . . were monitored by two or more experienced investigators." Having the sessions "monitored" does not necessarily mean that the workouts were supervised in such a way that the single-set group who was asked to train "to failure" actually did so.
- The varied multiple-set group did 3 x 5 for five weeks (2, 3, 8, 9 and 12), 3 x 3 for four weeks (4, 5, 10 and 13) and 3 x 2 for one week (14) while the one-set and three-set groups never did less than eight repetitions. The fact that the varied multiple-set group did ten weeks of training using lower repetitions — more than 70% of the total training period doing sets of 5 repetitions or less — certainly favored them when it came to tests of a 1-RM squat. This extremely poor experimental design stacked the deck against the single-set and multiple-set groups.
- The fact that the target weights for the multiple-set groups was set by the researchers undoubtedly placed the single-set group at a huge disadvantage. This selective favoritism toward the multiple-set groups is additional evidence of researcher bias.

Some comments about the test results and conclusions:

- The researchers found no significant changes in body mass or body composition over time or between groups ($p<0.05$). If body weight and body composition changes are to be evaluated, the experimental groups should receive direction in the area of caloric intake and expenditure. Ideally, these nutritional variables should be controlled.
- Though not statistically significant, the single-set group decreased their body-fat percentage from 13.6 to 13.5; the multiple-set group and the varied multiple-set group actually increased their body-fat percentage from 13.1 to 13.3 and from 15.6 to 15.8, respectively. Though not statisti-

cally significant, the single-set group increased their lean body mass from 67.5 to 67.9 kilograms. After 14 weeks of training, the varied multiple-set group did not show any change in lean body mass. During the last eight weeks of training, however, the varied multiple-set group actually lost lean body mass (0.3 kilograms). During the last eight weeks of training, the single-set group increased their body mass by 0.3 kilograms while the varied multiple-set group increased their body mass by 0.2 kilograms. To summarize the changes in body mass and body composition of the varied multiple-set group during the last eight weeks of training: They increased their body mass by 0.2 kilograms which was a result of an increase of 0.5 kilograms of body fat and a decrease of 0.3 kilograms of lean body mass.

- After 14 weeks, all three groups showed significant increases in their 1-RM squat, squat per kilogram of body mass and squat per kilogram of lean body mass. The multiple-set groups experienced significantly greater increases in their 1-RM squat compared to the single-set group. According to the researchers, the subjects were actually tested in the 1-RM "parallel squat." Further, they stated that the 1-RM parallel squat "was measured using the methods outlined" in a book by Stone and O'Bryant (1987). This book gives a "checklist" for the parallel squat in which a maximum total of 20 points is given for the execution of four techniques including the "starting stance," "down phase," "up phase" and "ending position." The book also describes a "parallel back squat" as when "the tops of the thighs are below parallel with the platform." There were no data reported in this study that indicated the "checklist" described by Stone and O'Bryant (1987) was used during the testing of the 1-RM squat. Furthermore, there was no indication as to how the researchers managed to be consistent in their judgment of whether or not "the tops of the thighs [were] below parallel" during the testing of the 1-RM squat. Needless to say, anyone who has ever competed in the sport of powerlifting knows that judging a squat to be "parallel" is highly subjective and can be extremely inconsistent — even by experienced judges who have attained national certifications. For this reason, any claims, comments or conclusions made by the researchers that are related to the test results of the squat — including 1-RM strength, squat per kilogram of body mass and squat per kilogram of lean body mass — are highly questionable.
- The groups were also compared in "volume load" (repetitions x mass lifted), "intensity" (average mass lifted) and "relative intensity" (percentage of initial 1-RM) of the target sets. The researchers calculated these rather bizarre measures from training logs that were kept by the subjects. Information provided by the subjects rather than the researchers contaminates the scientific purity of the data.
- The researchers stated that "The use of multiple sets in resistance exercise has been shown to produce superior maximum strength gains." As support for their statement, the researchers cited two references: Berger (1962) and Stowers and his co-workers (1983). The design flaws in the Berger study has been discussed previously; the design flaws of the Stowers study will be examined in the next issue of this newsletter. For now, however, the study did not show — as the researchers claim — that the use of multiple sets "produce[s] superior maximum strength gains." In truth, the Stowers study showed no significant differences between the one-set group and the three-set group in the 1-RM bench press and 1-RM squat after seven weeks of training.
- The researchers noted that "studies with a periodized protocol have generally shown superior strength increases in various measures of hip and leg maximum strength using untrained . . . subjects." As support for their assertion, the researchers cited four references: The first was an unpublished doctoral dissertation previously co-authored by one of the researchers (O'Bryant); the second was previously co-authored by two of the researchers (O'Bryant and Stone); the third was, strangely, a position stand of a national organization previously authored by one of the researchers (Stone); and the fourth was previously co-authored by one of the researchers (Stone).

IN PART II

The second part of this investigation into flaws in research design and interpretation will examine two additional studies: Stowers and his co-workers (1983) and Kraemer and his associates (1995).

References:

Berger, R. A. 1962. Effect of varied weight training programs on strength. Research Quarterly 33 (2): 168-181.

Carpinelli, R. C. 1999. The multiple-set myth. In Maximize your training: insights from leading strength and fitness professionals, ed. M. Brzycki, 81-95. Lincolnwood, IL: Masters Press.

Carpinelli, R. C., and R. M. Otto. 1998. Strength training: single versus multiple sets. Sports Medicine 26 (2): 73-84.

Kramer, J. B., M. H. Stone, H. S. O'Bryant, M. S. Conley, R. L. Johnson, D. C. Nieman, D. R. Honeycutt and T. P. Hoke. 1997. Effects of single vs multiple sets of weight training: impact of volume, intensity, and variation. Journal of Strength and Conditioning Research 11 (3): 143-147.

Stone, M. H., and H. S. O'Bryant. 1987. Weight training: a scientific approach. Minneapolis: Burgess.

Stowers, T., J. McMillan, D. Scala, V. Davis, G. D. Wilson and M. H. Stone. 1983. The short term effects of three different strength-power training methods. NSCA Journal 5 (3): 24-27.