Starting Strength

Training Optimality
A Study Proposal from an Economics Perspective

by

Nicholas M. Racculia

"The real question in any training discussion is not what works. Rather, you should ask, 'What is optimal?'" – Mike Zourdos

In finance, there is a search for an efficient investment frontier, in which an investor maximizes expected returns (which investors like) for any given level of risk (which keeps investors up at night). Given individual risk tolerances, and a few other factors like age, investment professionals discover an optimal collection of securities (bonds, stocks, etc) for each investor. This collection of securities is known as an investor's portfolio.

Training is a portfolio of exercises into which a trainee invests effort over time. The returns generated from these expenditures are increased fitness levels. They can be expressed quantitatively as strength, endurance, power, speed increases and by other less quantifiable markers – looking better, feeling better, etc. Whether a trainee invests for short term performance (a high school sport) or long-term health (our "Physical 401k"), a rational search for optimality is useful given that effort, time and money are all scarce resources.

What is optimal? For the vast majority of novices, there should be an objectively verifiable optimal training program – a first best solution to programming. A program's optimality should maximize efficiency and minimize the risk of injury based on the context of the trainee's situation. Exercise progression can have quantified rankings – e.g., Subject A's increase in his squat 1RM using Program X was 20% greater than Subject B's increase using Program Y – while training programs should have ordered ranking, e.g., Program X is better/worse than Program Y. Coaches discover what is effective and what should be left behind.

Efficient training maximizes progress per hour trained since only a limited number of hours can be typically devoted to fitness. The progress created by training is hampered by injuries. Think of the net effect as workout to workout profits (progress minus injuries). Repeated maximal short-term physical profits maximize long term physical wealth. An optimal program maximizes these profits within the context of long-term health.

When you invest your money, would you be satisfied with a 1% return? No. You would want to get the most out of the money that you can, since you only have a short lifetime in which to invest it. Training is no different: we want the highest return we can get in our physical 401K. These short-term

profits must be allowed to accrue, otherwise acute injury or the third stage of Seyle's general adaption syndrome [1] destroys progress.

The motivation for discovering optimal programming is rational. Though our life span is the longest thing we experience for sure, it is relatively short in terms of the limited number of training sessions available. We want to get the most out of each training session. Getting strong or increasing power or endurance should happen as efficiently and safely as possible. On the one side of the life cycle, efficient programming benefits athletic programs in high schools and colleges where there is a fairly limited amount of time to build strong athletes. Optimal programming minimizes downtime from injuries.

On the other side, aging populations face metabolic syndrome, osteopenia, sarcopenia and all sorts of other horrific conditions. An optimal training program enables grandma to independently get to her bridge tournament and helps prevent a broken hip if she slips on the ice on her way.

The compression of morbidity during times of rising healthcare costs alone is substantial motivation to search for optimality. Jonathon Sullivan illustrates this most convincingly: "Instead of slowly getting weaker and sicker and circling the drain in a protracted, painful descent that can take hellish years or even decades, we can squeeze our dying into a tiny sliver of our life cycle. Instead of slowly dwindling into an atrophic puddle of sick fat, our death can be like a failed last rep at the end of a final set of heavy squats. We can remain strong and vital well into our last years, before succumbing rapidly to whatever kills us. Strong to the end [2]."

It has been argued convincingly that strength is the most useful fitness adaptation [3]. Though my own personal bias agrees, this optimality framework should be adaptable to other quantifiable aspects of fitness as well. Fitness can be expressed in multiple ways. Jim Cawley has described ten general aspects of fitness [4]:

Cardiovascular/respiratory endurance: The ability of body systems to gather, process, and deliver oxygen.

Stamina: The ability of body systems to process, deliver, store, and utilize energy.

Strength: The ability of a muscular unit, or combination of muscular units, to apply force.

Flexibility: The ability to maximize the range of motion at a given joint.

Power: The ability of a muscular unit, or combination of muscular units, to apply maximum force in minimum time.

Speed: The ability to minimize the time cycle of a repeated movement.

Coordination: The ability to combine several distinct movement patterns into a singular distinct movement.

Agility: The ability to minimize transition time from one movement pattern to another.

Balance: The ability to control the placement of the body's center of gravity in relation to its support base.

Accuracy: The ability to control movement in a given direction or at a given intensity.

Program comparisons are not new in the literature. For the treatment of patients with coronary artery disease, Warburton et al. [5] find that interval training is more effective than traditional low intensity

cardio. The health benefits of high intensity training surpass those of moderate intensity according to a study by Wisloff et al. [6]. Tjonna et al. [7] also indicate increased benefits from high intensity versus moderate intensity for patients expressing metabolic syndrome.

Performance comparisons are also abundant in the literature. For example, Evertsen et al. [8] demonstrates that interval training may be more effective for a variety of biochemical changes that affect cross-country skiers. Storen et al. [9] show the benefits to elite cyclists from high intensity interval training exceed that of low intensity steady state training. Wilson et al. [10] find that training for multiple aspects of fitness tends to be detrimental to optimal gains as multiple training modalities compete for metabolic resources.

Most of these studies, while useful, are designed to test a specific question rather than the more general "how should I spend my time in the gym?" It would be useful to have some framework that can allow trainees to rank programs claiming to increase some aspect of fitness against other programs in a more generalized way.

There may be one such study already. Rhea et al. [11], a meta-analysis, combines a volume of previous literature and draws conclusions about optimal intensity, volume and frequency. One of the conclusions is suggestive: the study states that 60% of a one-rep max (1RM) is the best way to elicit maximal strength increases for the untrained [12].

The analysis seems flawed for a number of reasons. The authors define "untrained" by the length of time spent in the gym (1 year or less). There are many efficiently programmed trainees who achieve intermediate training status before a year's time. Likewise, there is many a cardio bunny who after spending years on the elliptical would respond like a novice to a proper strength program. A more precise metric for judging level of progression, such as time to recovery, is not present in the paper.

Furthermore, since strength is defined as the ability to generate force against an external object, how does training at 60% of maximal strength increase maximal strength most efficiently [13]? Such a low percentage of a 1RM maximizes local muscular stamina gains rather strength gains, and that only with sufficient training volume. Finally, it is important to note that only 33 out of 1,063 observations in the study (or 3%) trained at 60% 1RM (rather than 70%, 80%, etc). This suggests that the sorting mechanism itself precludes us from drawing a generalized conclusion.

Moreover, pre-testing a 1RM for a novice is essentially impossible (they are just learning the movement patterns), entirely fruitless (since a rank novice cannot usher forth the force necessary to push through a true 1RM) and ultimately pointless (since the attempt itself will cause an adaptation rendering a new 1RM). Attempting to calculate a 1RM using generic tables is equally unhelpful. Reason and experience (and three other studies by Hoeger et al. [14], Hoeger et al. [15], LeSuer et al. [16]) show that these tables should be specific not only to the exercise (why would we use the same equation for a squat and a bench press, and why do these equations usually assume linearity) but to the person as well. Confounding effects like gender, anthropometry and willpower render such a pursuit useful only for impressing potential lovers and, more importantly, fellow gymrats.

Also, though the statistical analysis employed seems reasonable (effect sizes with a fairly large data set), the input data may be suspect, which calls into question the effect sizes. Some randomly drawn examples (from the studies on untrained individuals only) from the meta-analysis include faulty data.

For example, Chilibek et al. [17] use the leg press, the bench press and the arm curl (with knee extension, knee flexion and the lat pulldown machines included in the training but not in the testing for "balance") as a proxy for total-body training during a 10-week training program. The movement pattern for each exercise is not clearly defined. The pre- and post-1RM were used to show changes

in strength, which, as described earlier is entirely unhelpful. Finally, a curious result, given that all participants were female – there was a 72.5% increase in curl strength and only a 21% increase in leg press strength. This input does not bode well for Rhea et al.

Another randomly selected contributor to this meta-analysis is Lemmer et al. [18]. The study uses younger (~25 yrs) and older (~69 yrs) volunteers (variety is useful for making general predictions) to study differential strength increases and detraining effects by using a knee extension machine on only one leg. It does not make sense to include this study in the meta-analysis for a multitude of reasons, not the least of which is that adaptation driven by a single leg extension is significantly different from the adaptation driven by a thoughtful total-body program.[19]

If a study on one-legged extensions was the only outlier, then perhaps the input data aren't too flawed. The paper also cites Zmierski et al. [20] and its work on scapular strengthening. Determining "if isokinetic strengthening of the scapular adductors while horizontally abducting the shoulder is more effective than strengthening the scapular adductors while extending the shoulder" is no doubt an interesting study, but is it appropriate to include in a meta-analysis on optimal strength dosing?

When suspect data arrives at an untenable conclusion, a new approach may be necessary [21].

The Model

First, this model applies only to untrained individuals. It becomes much more difficult to predict optimal programming for more advanced lifters as diminishing marginal returns make progress a highly individualized effort. Thus, generalized optimal training will only apply to a novice.

Novice: A novice is defined as a trainee who can adapt from a disruption in homeostasis (i.e. training) in 48 to 72 hours.

Non-Novice: A trainee who requires a disruption from homeostasis with an intensity level high enough to require more than 72 hours of recovery time for adaptation to occur.

Optimality is a nebulous term. Let's defined some terms more carefully, since the central question is optimality:

Training Optimality: A program exhibits Training Optimality if, for any novice progression, the program is quantifiable and strictly efficient within one aspect of fitness and non-detrimental to the other aspects of fitness. A program that is both efficient and non-detrimental (defined below) can be called optimal for a novice trainee; see Figure 1.
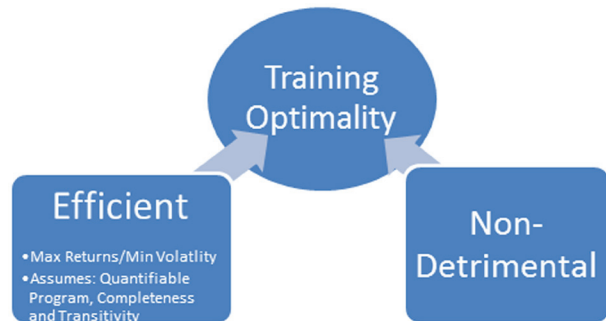

Figure 1.

Efficient: A program exhibits Efficiency if there is no known program which increases one aspect of fitness at a faster rate. Here is the central testable quantity: efficiency. The ability to define efficiency is based on at least three assumptions, that programs are quantifiable and exhibit completeness and transitivity.

A program cannot be measured if it is not quantifiable. So let's define quantifiable:

Quantifiable: An exercise is quantifiable if progress can be measured, tracked and compared.

For example, suppose a coach has established good form for a trainee's squat and has just finished the first work sets at 135 lb. If in one month that trainee's squat goes from 135lbs for sets of 5 across to 195 lbs for sets of 5 across, the trainee can move approximately 60 lbs more [22]. A program is the aggregation of that quantifiable progress. The quantification must be comparable across time and against other programs, and is necessary to search for optimality. If progress cannot be tracked over time, how can progress be determined?

Quantification should be limited to activities (weight lifted, sprint time, distance run, etc.) rather than other metrics (body weight, belt size, feelings of exhaustion, soreness, etc.). Activities can yield universally comparable results while these other metrics are less comparable from person to person. A bench press, done correctly, and absent severe differences in height above sea level, generates the same external force for all participants.

Efficiency, if quantifiable, needs two additional assumptions: Completeness and Transitivity.

Completeness: Completeness is the idea that any two programs can be compared. Completeness has number of implications. First, this framework only applies to systemic programming, programming that effects the entire system. Within the context of strength it is rather silly to claim a program makes you stronger if only one part of your body (for example, "upper body") becomes stronger [23]. This is not to suggest that every one of the 642 skeletal muscles must be worked to be considered systemic, just the vast majority. I mean, can you really linearly program the strengthening of the occipitofrontalis? (Perhaps, if the trainee with a serious program spends enough time watching foolishness in the gym…)

Second, a program must have precise descriptions of its constituent exercises. It is unnecessary for all programs to contain identical exercises, since the post-test will have all participants performing (with randomized order) the 1RM for each exercise in each program. In other words, if Program X uses squats (as defined by the Starting Strength method) and Program Y uses quarter squats, all participants will post-test their 1RM for both exercises (see below for a pilot setup). All that is required is that each exercise is given a precise description, since, squat ≠ quarter squat, no matter what your high school football coach tells you.

Strength: the ability to produce force against an external object. If two programs are systemic in nature and their constituent exercises are precisely defined, it stands to reason that they can be roughly sorted into an ordinal ranking (better, worse, same) for the entire system. If not, and, for example, Program X displays significant dominance in, say upper body strength increases but significant weakness in lower body strength increases compared to Program Y, then it may be useful to explore combining the superior aspects of each program.

Transitivity: If Program X exhibits efficiency (as we are defining it) over Program Y and Program Y exhibits efficiency over Program Z, then Program X is more efficient than Program Z. The inclusion of transitivity allows the researcher to rank multiple programs from best to worst.

Strict Efficiency: Program X exhibits Strict Efficiency over Program Y if all exercises within Program X increase at least as fast as comparable exercises in Program Y, with at least one exercise from Program X increasing faster than Program Y.

An implication of efficiency is the lack of, or minimization of, deviations from predicted outcomes. If a program makes aggressive claims as to expected improvements and the majority of participants are not hitting those goals even when following the program, then the efficiency is called into question.

The measure of this uncertainty is called volatility.

> *Volatility:* the sum of deviations from some expected, quantifiable unit of progress within the aspect of fitness (e.g., strength = 1RM).

Total volatility, or deviation, comes from two sources – (a) you don't do the program as described or (b) you follow the program but it fails to accurately predict progress. Let's call the former "idiosyncratic volatility" and the latter "programmatic volatility." Total deviation is the sum of programmatic volatility and idiosyncratic volatility. For simplicity, assume that programmatic volatility and idiosyncratic volatility are independent and the presence of one doesn't correlate with the presence of the other [24].

> *Programmatic Volatility:* Program volatility represents the sum of the deviations from predicted outcomes in a program's design. A training protocol with a high degree of volatility built into its structure should necessarily have poorer performance results than one which does not. If trainees follow all aspects of the program as prescribed (including diet and sleep recommendations) and their increases deviate from the claimed design, then the method exhibits significant program volatility.

> *Idiosyncratic Volatility:* Idiosyncratic volatility represents the sum of the deviations from expected outcomes due to not following the program's design. Idiosyncratic deviations includes things like not sleeping enough, adding additional workouts or exercises, not eating enough (or too much) and not maintaining proper form. They may also include external factors like having a piano fall on your head. The presumption here is that a trainee, before the fact, will do the program as prescribed.

Idiosyncratic volatility should have an expected value of zero for any given lifter, though a self-selection bias may be present. Some programs might attract more serious trainees than other programs. Some programs may attract more risk-tolerant trainees than others. This suggests program and idiosyncratic volatility may very well not be independent. Thoughtful data collection can attempt to discern between idiosyncratic deviations and programmatic deviations.

Total volatility is then the sum of these two sources of variation. Suppose a program claims linear progression in the squat. See Table 1 for what should be a nine week progression. Missed reps would necessitate the trainee repeating the same weight in the following session. The sum of the deviations represents the total volatility. In this particular example, our trainee increased strength by 25 lbs with a 15 lb deviation from the predicted strength path. The training log makes note that twice our trainee failed to act according to the program's prescription. Idiosyncratic volatility contributes about 10 lbs of the deviation while the remaining 5 lbs of deviation can be attributed to programmatic volatility.

Training Session	Predicted Program Gains	Actual Weight Lifted
1	135	135
2	140	140
3	145	145
4	150	145 (slept poorly)
5	155	150
6	160	155
7	165	155 (ate poorly)
8	170	155
9	175	160

Table 1. Predicted Gains and Volatility.

A program can be efficient, but if it causes a significant decrease in other aspects of fitness, then, for a novice, it would not be optimal.

Detrimental: A program is detrimental for a novice if increases in one aspect of fitness (e.g., strength) coincide with materially significant decreases in one or more aspects of fitness (e.g., losses in speed and endurance). A program that does see decreases in one or more aspects of fitness would not be classified as "non-detrimental." This definition does not apply to trainees no longer in the novice phase of training.

A First Test

Here is a testable hypothesis. Start with a robust pool of self-reported evidence on strength gains using the Starting Strength program. Go a step further and state that as we get stronger, novices who have completed the program do not experience reductions in speed, endurance, agility and all the other aspects of fitness.

Hypothesis: The Starting Strength Novice Program is Optimal with respect to Strength acquisition.

Best programs should follow from results. The goal of each experiment is to reject the hypothesis. Failing to reject the hypothesis does not prove that Starting Strength is optimal, but it does show that some other program is not optimal. With time, the fitness industry will phase out the junk heap of suboptimal programs.

Corollary: The Starting Strength Novice Program is preferred to the ACSM strength training protocol.

Each program is defined as follows:

Starting Strength Novice Program: The Starting Strength protocol centers around five barbell lifts (squats, deadlifts, presses, bench presses and power cleans) and chin-ups. Training occurs three times per week on non-consecutive days alternating A and B workouts, see Table 2, with chin-ups added. After appropriate warm-up sets, the volume is 3 work sets across of 5 reps for squats, bench press and press, 1 set of 5 reps for deadlifts, and 5 sets of 3 reps for power clean. The load follows a quasi-linear progression [25] with weight added to the bar for all lifts for all training sessions for as long as possible.

Day A	Day B
Squat	Squat
Bench Press/ Press	Press/Bench Press
Deadlift	Power Clean

Table 2. Starting Strength Programming

ACSM Strength Training Pr otocol: The ACSM offers multiple strength training protocols on its website. Here are guidelines for their general strength program:

> "The American College of Sports Medicine (ACSM) recommends that a strength program should be performed a minimum of two non-consecutive days each week, with one set of 8 to 12 repetitions for healthy adults or 10 to 15 repetitions for older and frail individuals. Eight to 10 exercises should be performed that target the major muscle groups." [26]

They provide "examples of typical resistance exercises," detailed in Table 3.

The ACSM recommends that the optimal load to build "Muscular Strength" [27] is 60-70% of 1RM with a volume of 1-3 sets. Progression of loading using the ACSM protocol is described thus:

"a 2-10% increase in the load be applied when the individual can comfortably perform the current workload for one to two repetitions over the desired number on two consecutive training sessions."

	Free-Weight	Machine-Based	Body Weight
Chest	Supine Bench Press	Seated Chest Press	Push-ups
Back	Bent-Over Barbell Rows	Lat Pulldown	Pull-ups
Shoulders	Dumbbell Lateral Raise	Shoulder Press	Arm Circles
Biceps	Barbell/Dumbbell Curls	Cable Curls	Reverse Grip Pull-ups
Triceps	Dumbbell Kickbacks	Pressdowns	Dips
Abdomen	Weighted Crunches	Seated "Abs" Machine	Crunches, Prone Planks
Quadriceps	Back Squats	Leg Extensions	Body Weight Lunges
Hamstrings	Stiff-leg Deadlifts	Leg Curls	Hip-ups

Table 3. ACSM Typical Resistance Exercises

Pilot Setup

Week 1: Prior to group assignment, teach all of the test lifts, establish baseline strength, speed, endurance, etc. for each subject. Then randomly assign to SS and ACSM groups. Teach the rest of the program and assign caloric intake, recommend sleep levels.

Weeks 2-13: Program progress tracked; calories and sleep tracked; injuries noted; deviations from expectations classified as programmatic or idiosyncratic [28].

Week 14: Re-Introduce all Test lifts. Test 1RM for each test lift, which is now possible given that all trainees have some "time under tension," to borrow a phrase, and can safely and effectively exert enough force to generate a truer 1RM. As long as the sample size is large enough, pre-test 1RMs are unnecessary since it is reasonable to assume similar starting strengths for two large enough untrained populations. The difference in ending 1RM is all that will be tested.

   Test Efficiency: Test for statistically significant [29] difference for each of the test lifts. Compare results.

   Test Non-Detrimentality: Tests for other aspects of fitness – compare pre/post for each program. Test for materially significant declines.

Life is short. Perfect optimal training may never be discovered, but aiming for predictable and maximal efficiency ought to be the goal of any coach and any serious trainee. The fields of economics and finance have long used terms like efficiency and optimality. Perhaps it is time we introduce similar concepts into physical culture.

Nicholas M. Racculia is an Assistant Professor of Finance at Saint Vincent College. When not torturing students by making them occasionally learn finance, he enjoys research on venture capital.

He is a graduate of Princeton University (PhD, Economics) and Saint Vincent College. He earned his Starting Strength Coach credential December 2012.

Starting Strength : Resources : Articles : Forums : Discuss This Article

References & Notes

1 Selye, Hans. A Syndrome Produced by Diverse Nocuous Agents. *Nature* 1936; 138: 32.

2 Sullivan, Jonathon. Barbell Training is Big Medicine.

3 Wolf, Michael. Strength & Barbells: The Foundations of Fitness.

4 Crawley, Jim. Ten Domains of Fitness. .

5 Darren E.R. Warburton, Donald C. McKenzie, Mark J. Haykowsky, Arlana Taylor, Paula Shoemaker, Andrew P. Ignaszewski, and Sammy Y. Chan. Effectiveness of High-Intensity Interval Training for the Rehabilitation of Patients With Coronary Artery Disease. *The American Journal of Cardiology* 2005; Vol. 95.

6 Ulrik Wisloff, Oyvind Ellingsen and Ole J. Kemi. High-Intensity Interval Training to Maximize Cardiac Benefits of Exercise Training? *Exercise and Sports Science Review* 2009

7 Arnt Erik Tjønna, Sang Jun Lee, Øivind Rognmo, Tomas O. Stølen, Anja Bye, Per Magnus Haram, Jan Pål Loennechen, Qusai Y. Al-Share, Eirik Skogvoll, Stig A. Slørdahl, Ole J. Kemi, Sonia M. Najjar, Ulrik Wisløff. Aerobic Interval Training Versus Continuous Moderate Exercise as a Treatment for the Metabolic Syndrome. *Circulation* 2008.

8 Evertsen, F., J.I Medbo and A. Bonen. Effect of training intensity on muscle lactate transporters and lactate threshold of cross-country skiers. ACTA Physio Scand 2001; 173: 195-205.

9 Oyvind Storen, Solfrid Bratland-Sanda, Marius Haave and Jan Helgerud. Improved V0(2) Max and Time Trial Performance with more High Aerobic Intensity Interval Training and Reduced Training Volume: A Case Study on an Elite National Cyclist. *Journal of Strength and Conditioning Research* 2012; 26: 2705-2711.

10 Wilson, Jacob, P. Marin, M. Rhea, S. Wilson, J. Loenneke and J. Anderson. Concurrent Training: A Meta-Analysis Examining Interference of Aerobic and Resistance Exercises. *Journal of Strength and Conditioning Review* 2012; 26.8: 2293-2307.

11 Rhea, Matthew, Brent Alvar, Lee Burkett and Stephen Ball. A Meta-analysis to Determine the Dose-Response for Strength Development. *Med. Sci. Sports Exerc* 2003; 35-3: 456–464.

12 The meta-analysis separates studies into trained individuals and untrained individuals. I safely conclude that all novices (as a function of recovery time) must be a subset of "untrained" as the paper defines it.

13 Perspective may help –This suggests that a rank novice (who happens to be able to squat 60lbs once) should be training with work sets using a Rogue Bella Bar, ONLY.

14 Hoeger, W., S. Barette, D. Hale and D Hopkins. Relationship Between Repetitions and Selected Percentages of One Repetition Maximum. *Journal of Applied Sports Science Research* 1987; 1-1:11-13.

15 Hoeger, W., S. Barette, D. Hale and D Hopkins. Relationship Between Repetitions and Selected Percentages of One Repetition Maximum: A Comparison between Untrained and Trained Males and Females. *Journal of Applied Sports Science Research* 1990; 4-2:47-54.

16 LeSuer, D., J. McCormick, J. Mayhew, R. Wasserstein and M. Arnold. The Accuracy of Prediction Equations for Estimating 1-RM Performance in the Bench Press, Squat and Deadlift 1997; 11-4:211-213.

17 Chilibeck, P.S., A.W. Calder, D.G. Sale and C.E. Webber. A comparison of strength and muscle mass increases during resistance training in young women. *Eur. J. Appl. Physiol. Occup. Physiol* 1998; 77:170–175.

18 Lemmer, J.T., D.E. Hurlbut, G.F. Martel, et al. Age and Gender Responses to Strength Training and Detraining. *Med. Sci. Sports Exerc* 2000; 32:1505–1512.

19 One other curiosity in this study—subjects were given 60 seconds of rest between 1RM attempts. Anyone who has ever lifted anything close to a 1RM knows that 60 seconds between attempts is not quite enough time.

20 Zmierski, T., S. Kegerreis and J. Scarpaci. Scapular Muscle Strengthening. *Journal of Sport Rehabilitation* 1995; 4:244- 252.

21 Rippetoe, Mark. The Problem with Exercise Science.

22 Whether this is a result of increased neuromuscular efficiency or myofibrillar hypertrophy is really irrelevant.

23 Wells, Damon. Adaptation: Period, Persistence and Prioritization.

24 This assumption may not be realistic, but it is a starting point.

25 The weight will increase at a decreasing rate towards the end of the novice progression (e.g., 5 pound jumps in the squat eventually turn into 2.5 pound jumps).

26 Esco, Michael. Resistance Training for Health and Fitness.

27 Are there other types of strength besides muscular strength that the ACSM can train??

28 Though a novice progression may take longer than 3 months depending on individual genetics, capping it at three months should ensure that the vast majority of participants have not left the novice phase but have progressed long enough to have progress that can be compared.

29 Though maybe "material difference" might be more useful?