



Best Practices in the Design of Aerobic Training Programs

*Jason Jameson, Ph.D.
Ross R. Vickers Jr., Ph.D.*



Naval Health Research Center

Document No. 10-12

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government. Approved for public release; distribution is unlimited.

This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.

*Naval Health Research Center
140 Sylvester Rd.
San Diego, California 92106-3521*

Best Practices in the Design of Aerobic Training Programs

Jason Jameson
Ross R. Vickers, Jr.

Warfighter Performance Department
Naval Health Research Center
140 Sylvester Road
San Diego, CA 92106-3521

e-mail: jason.jameson@med.navy.mil

This represents report 10-12 and was supported by the U.S. Department of Defense under work unit 60704. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government. Approved for public release; distribution is unlimited. This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.

Abstract

Successful military operations depend on the aerobic fitness of military personnel. Training programs that tax the cardiorespiratory system are known to increase aerobic fitness, and program design choices influence the magnitude of these gains. This review attempted to identify design choices that could be considered best practices. A best practice is a design option (such as training at an intensity of 90% of one's maximum heart rate) that produces significantly better results than any other option (e.g., training at 60%). To this end, this review employed meta-analytic techniques to synthesize studies that investigated the design options that determine aerobic fitness. To ensure sensitive assessments of program design effects, statistical procedures adjusted for the repeated measures structure of the study designs. Unfit individuals benefitted much more from training than fit individuals. Gender and age were not influential moderators. Regarding program design options, the intensity of a training program, the duration of a training session, the frequency of training per week, and the length of a training program were all significant moderators. However, with the exception of training intensity, post hoc comparisons generally showed that no single design option was significantly better than all others. The available evidence may rule out some design choices, but it is too limited to identify best practices.

Summary

Successful military operations depend on the aerobic fitness of military personnel. Training programs that tax the cardiorespiratory system are known to increase aerobic fitness, and program design choices influence the magnitude of these gains. This review attempted to identify design choices that could be considered best practices.

Issue

Because successful military operations depend on the aerobic fitness of military personnel, effort must be devoted to the design of optimal training programs. Given the many design options available, to what extent does the current literature on aerobic training prioritize some options over others?

Objective

The purpose of this meta-analysis is to integrate results across several studies to determine the effects of many factors on aerobic fitness—such as those that relate to the training program (e.g., frequency, intensity, duration, and mode of exercise), in addition to the program participants (e.g., initial fitness level).

Approach

Statistics describing the effects of training on aerobic fitness were extracted from journal articles. Every study included in this review employed a pretest-posttest design. To determine the effect size due to training, estimates of the training response were adjusted for the type of research design. Meta-regression models evaluated potential moderator variables (a demographic variable or program element that might account for variation in the overall effect size due to training). Best practices were evaluated through post hoc comparisons between different levels of each moderator variable.

Results

Unfit individuals benefitted more from training than fit individuals. Gender and age were not influential moderators. The intensity of a training program, the duration of a training session, the frequency of training per week, and the length of a training program were all significant moderators. However, with the exception of training intensity, statistical tests generally showed that no single design option was better than all others. The available evidence may rule out some design choices, but it is too limited to identify best practices.

Reviews of the aerobic training literature have shown that training increases aerobic fitness (Londoree, 1997; Samitz & Bachl, 1991; Wenger & Bell, 1986). The same reviews have established that aerobic fitness is influenced by training program design. With the effectiveness of aerobic training well established, attention shifts to the question of whether it is possible to identify best practices for aerobic training. A best practice is a specific program design option that is superior to all other possible choices for that program design facet. For example, the intensity of each training session (measured, for instance, in terms of the percentage of an individual's maximum heart rate) is a program design facet. If the cumulative research record indicated that an intensity equal to 90% of one's maximum heart rate produced significantly better results than any other choice for this facet, then 90% of the maximum heart rate would be a best practice.

This review attempted to identify best practices based on the available evidence, by focusing on maximal oxygen uptake (i.e., $\dot{V}_{O_{2max}}$) as the key dependent variable of interest. This measure is typically considered the gold standard for indexing cardiorespiratory fitness. Previous reviews have attempted to estimate the effects of several program design facets on aerobic fitness, such as training intensity or session duration. In addition, these meta-analyses may be viewed as also studying best practices, but they have done so only indirectly, and have not formally attempted to identify differences in design facets as best practices. This meta-analysis explores the contribution of these factors within the context of formally identifying best practices.

The second difference between this review and prior reviews involved the treatment of statistical issues. One set of issues derived from the repeated measures structure of the evidence. Aerobic fitness studies routinely employ repeated measures research designs. Tests of aerobic capacity are administered before the training program begins, and again after the program has been completed. The difference between the pre and post training scores is the basis for estimating the effect size (ES) for the training program. Steps must be taken to adjust for repeated measures experimental designs when estimating a study's ES (Morris & DeShon, 2002).

Another statistical issue derived directly from the current interest in identifying best practices. It is not enough to demonstrate that program design choices affect the size of the training response. Analysis of variance (ANOVA) tests have been used to test the hypothesis that the effects of different design choices are the same. Rejecting this null hypothesis has only indicated that some options differ from other options. It is not enough to know that differences between options exist. The existence of differences does not guarantee the existence of a best choice. Thus, a significant ANOVA must be followed by analyses that evaluate differences between specific program design options. This review employed post hoc comparisons to determine whether the design option that produced the largest ES was truly a best practice.

This review attempts to identify best practices for several design facets of aerobic training programs. Statistical methods are introduced to analyze the repeated measures structure of the data and the need for post hoc comparisons to determine whether a significant moderator effect truly identifies a best practice. As a result, this review provides a different perspective on the available evidence. In particular, this review attempts to formally identify best practices for the program design facets of training intensity, session duration, the number of training sessions per week, the length of the training program, the type of exercise (e.g., cycling or running), and the type of training program (interval or continuous). The search for best practices also considered initial

fitness level as a key demographic variable that might influence the impact of different program design choices.

Methods

Literature Search Procedures

The initial search of the literature centered on the PubMed database. The search terms included various combinations of the following: “training,” “aerobic training,” “aerobic fitness,” “cardiorespiratory fitness,” “cardiorespiratory training,” “cardiovascular training,” “maximal oxygen consumption,” “VO₂max,” and “functional capacity.” The search produced a list of 6,099 candidate articles. This list was narrowed down by excluding studies with participant samples consisting of animals, patients, children, adolescents, and those who were obese or who were diabetic (the exclusionary criteria are discussed in greater detail below). These criteria generated a reduced list of 3,814 candidate articles, which was further reduced to 756 by requiring that only articles with experimental trials be included. The PubMed abstracts for the remaining 756 articles were reviewed to determine whether they met the inclusion criteria for this review. Articles were dropped at this point in the search only if the information in the abstract clearly indicated that the study failed to meet at least one of the criteria. In addition to the previously stated exclusionary criteria, subsequent screening required the studies to have some measure of $\dot{V}_{O_{2\max}}$, a specific aerobic training program, and to have no specialized respiratory treatment. These criteria reduced the list to 251 articles. In addition to the PubMed search, 150 articles that contributed data to previous aerobic fitness reviews were also examined (Londoree, 1997; Samitz & Bachl, 1991; Wenger & Bell, 1986).

The full texts of 401 articles that passed the initial screening process (251 from the PubMed search; 150 from previous reviews) were examined to determine whether the studies met the following criteria:

1. Study participants were required to be healthy. This criterion led to the exclusion of studies whose participants were hypertensive or diabetic (among other conditions). As a general rule, a study was excluded if the study participants were described as “patients.” However, studies of “overweight” individuals were accepted, so weight considerations eliminated only studies of individuals toward the upper end of the excess weight range. The objective in making these exclusions was to eliminate studies that might produce atypical effects because of limitations on the ability to perform training exercises, and/or that involved disease and metabolic processes that might modify the training response.
2. Study participants could be no younger than 16 nor older than 50 years of age. This criterion was intended to restrict the study samples to a population more similar to typical military personnel, in addition to minimizing the confounding of training effects with the effects of normal developmental processes.
3. Maximal oxygen uptake ($\dot{V}_{O_{2\max}}$) was expressed in milliliters of oxygen per kilogram per minute (i.e., ml·min⁻¹·kg⁻¹).

4. The study reported pre and post training measures of $\dot{V}_{O_{2\max}}$ and the standard deviations for those measurements. This information was the minimum required to compute ES when combined with assumptions about the magnitude of the pretest-posttest correlation (see Appendix A). $\dot{V}_{O_{2\max}}$ was measured more than twice in some studies. When this was the case, ES was computed using the initial and final measurements. Computing the effect for each phase of the training programs would have increased the complexity of the repeated measures problem. Therefore, ES always represented the final cumulative training impact.
5. The training program was endurance-based rather than resistance-based.
6. Study participants were given no medications (e.g., beta blockers, such as propranolol). However, for those studies that evaluated the effects of different medications on aerobic capacity, placebo groups, when they were reported, were included in the analysis.

The final database consisted of data from 181 studies that met the inclusion criteria. Control groups from those studies were excluded from the review, as they were independent groups that participated in no training program. With this restriction, 294 samples provided sufficient data to be included in this review. The cumulative sample size was 3,382 study participants.

Table 1
Sample Characteristics

	k	ΣN	Mean	SD	Minimum	Maximum
Age	290	3342	26.75	7.49	15.6	50.5
Height	197	2211	173.80	7.02	154.1	186.9
Weight	230	2588	72.63	11.62	48.9	162.0
Percent body fat	81	1026	20.65	6.20	8.2	36.0

Note. The statistics describe the population of study samples rather than a population of individuals. The data were not weighted for the computations that generated these descriptive statistics.

Demographic and Methodological Variables

Age, height, weight, percent body fat. Age, height, weight, and percent body fat were coded from descriptive statistics reported in the studies, and a summary of these data is provided in Table 1. Note that these statistics are based on sample means, and not on data from individuals.

Gender. For most studies, the samples were composed entirely of men or entirely of women. Some studies consisted of samples of both sexes, and a few provided no definite information regarding gender. To represent this variability, gender was coded as men, women, or men and women combined.

Age. For most studies, the average age of study participants was reported separately for each independent treatment group in the study. For other studies involving multiple independent groups, only the overall mean age was provided. When this was the case, the separate groups were assigned the overall means. An age range (e.g., 18 to 22 years) was another common reporting method. Finally, some studies did not report age directly, but provided age-related demographic

information (e.g., university students). Qualitative data were coded based on judgments of the age range that would be typical of the group described. Age was classified into four categories: younger than 20, 20-29, 30-39, and 40 and older.

Initial Fitness Level. The initial fitness levels of study participants were inferred from their overall initial $\dot{V}_{O_{2\max}}$ values. The initial coding employed seven categories: very poor, poor, fair, average, good, very good, and excellent. A study sample was classified into an initial fitness level category if the mean initial $\dot{V}_{O_{2\max}}$ value for the sample participants fell within a particular range (usually of 3 to 4 $\dot{V}_{O_{2\max}}$ units). The coding scheme incorporated gender and age differences. Men, in general, were associated with higher $\dot{V}_{O_{2\max}}$ values, and age was inversely related to aerobic capacity. The specific ranges of $\dot{V}_{O_{2\max}}$ values that served as the basis for coding are provided in Appendix C.

Program Design Facets

Program length. Program length was the number of weeks that the training program lasted.

Intensity. The studies varied in how training intensity was characterized. In most cases, intensity was defined in terms of the percentage of maximum heart rate, $\dot{V}_{O_{2\max}}$ percentage, or percentage of heart rate reserve. When a range of percentages was provided (e.g., 85%-95% of maximum heart rate), intensity was recorded by taking the midpoint of the range. When multiple ranges were provided, the average of the separate midpoints was recorded. Given the variation in how intensity was reported across studies, a common classification scheme was adopted, based on Heyward (2006). The scheme enables classification of distinct physiological measurements into 3 categories: (1) *moderate*, (2) *hard*, and (3) *very hard to maximal*. The full classification scheme is provided in Appendix C.

Duration of a training session. This variable refers to the duration of a single training session, measured in minutes. This variable was coded into five categories, based on 15 minute intervals: less than 15 minutes, 16-30, 31-45, 46-60, and 61 and greater.

Frequency. This variable refers to the number of times that study participants trained during a given week. The number of sessions per week was described in terms five categories: 1-2 sessions per week, 3, 4, 5, or 6 and greater.

Type of exercise. This variable refers to the type of exercise adopted during the training program. The type of exercise typically was either cycling (most often on a cycle ergometer), or running (including jogging or walking, either on a track or treadmill). In some cases, a training program incorporated both cycling and running, which was classified as a distinct category. Some studies included other kinds of exercises, such as tennis or cross-country skiing, but they occurred individually so infrequently that they were classified together as “other.”

Type of training program. This variable refers to whether the training program involved intermittent (i.e., work periods separated by rest periods) or continuous training.

Table 2 presents the distribution of the demographic variables and program training characteristics. For each variable, the table provides the corresponding number of samples and effect sizes, and number of participants summed across samples.

Table 2
Distribution of Demographic Variables and Program Training Characteristics

	No. of samples	No. of ESs	ΣN		No. of samples	No. of ESs	ΣN
	Gender				Duration (minutes per session)		
Men	86	181	2025	<15	10	13	180
Women	32	68	760	16-30	61	104	1221
Men and Women	22	39	541	31-45	65	82	966
				46-60	22	35	364
				> 61	14	19	256
	Age group				Frequency (sessions per week)		
< 20	19	29	307	1-2	15	20	253
20-29	117	181	1958	3	80	132	1629
30-39	29	52	576	4	39	58	620
> 40	22	28	501	5	30	47	524
				> 6	19	25	225
	Initial fitness				Program length (in weeks)		
Very poor	4	5	60	1-4	19	26	243
Poor	14	15	167	5-6	20	27	287
Fair	55	73	984	7-8	38	65	701
Average	79	109	1280	9-10	38	63	728
Good	34	49	482	11-13	29	51	577
Very good	19	25	249	> 14	37	60	827
Excellent	14	18	160				
	Intensity				Activity		
Moderate	13	14	126	cycling	67	92	943
Hard	133	195	2309	run/walk	88	152	1800
Very hard	30	42	482	both	10	12	117
Maximal	5	8	59	other	31	37	508

Analysis Procedures

Every study included in this review employed a pretest-posttest design. For this reason, methods described by Morris and DeShon (2002) were applied to compute appropriate ESs for repeated measures (simple ES_{RM} ; see Appendix A). Meta-regression models evaluated potential moderator variables. A moderator was a demographic variable or a program element that might account for variation in ES_{RM} . The meta-regression analyses applied Hedges and Olkin's (1985) general methods. These methods included weighted ANOVA and weighted linear regression. The weight variable was the inverse of the estimated variance for ES_{RM} . In certain cases, not all of the moderator variable levels contained comparable data, thus separate analyses were performed. However, when comparable data were available, analyses with combinations of independent variables were conducted.

Moderators were evaluated in two steps. The first step was an overall test for a moderator effect to determine whether the ES_{RM} differed significantly across the levels of the moderator variable. The second step was taken only if there was a statistically significant moderator effect. Post hoc comparisons were conducted to determine which groups differed significantly. The average ES_{RM} values for the moderator groups were ranked from largest to smallest. The group with the largest average ES_{RM} was adopted as the reference group. The first post hoc test compared the reference group to the group with the second largest average ES_{RM} . If these two groups differed significantly, the post-hoc comparisons stopped at this point. If the two groups did not differ significantly, the group with the third-largest average was compared to the reference group. The comparisons continued down the ranked-ordered moderator groups until a significant difference was found. The comparisons stopped at that point, and all remaining groups were classified as differing significantly from the reference group.

Some post hoc comparison procedures required multiple significance tests. Performing multiple significance tests increases the probability that at least one comparison would be statistically significant by chance alone. A Bonferroni significance criterion was adopted to fix the analysis-wide probability of error at 5% or less. The post hoc procedures involved $j - 1$ comparisons for a moderator with j levels. The Bonferroni criterion for each moderator was $p_{critical} = .05/(j-1)$.

The post hoc comparisons identified equivalence sets. These sets consisted of the design option with the largest average effect, plus the alternative options that produced effects that were not significantly different from this reference value. The sets were equivalent in the sense that the alternative options in the set could not be confidently classified as less effective than the optimum design option based on the available evidence.

Large samples can produce significant results even for trivial differences (Rosenthal & Rosnow, 1984). To avoid mistaking sample size for explanatory power, the Tucker-Lewis index (TLI; Tucker & Lewis, 1973) was adapted to provide an ES index for the moderator analyses. This index is the proportion of the greater-than-chance variation in ES_{RM} accounted for by a moderator or set of moderators (see Appendix B). Cohen's (1988) ES criteria were applied to characterize the TLI as indicating trivial, small, moderate, or large moderator effects.

Funnel plots were constructed to evaluate the potential effects of publication bias (Light & Pillemer, 1984). The file-drawer problem was not examined because both the typical ES and the total number of studies were large. Under those circumstances, Rosenthal's (1979) file drawer criterion would almost certainly be satisfied.

Analyses were carried out with the SPSS-PC, Version 17, computer program (SPSS, Inc., Chicago, IL) and R, package version 1.5.2. (R Development Core Team, Vienna, Austria).

Results

Program Length Effect

ES_{RM} generally increased with program length. Preliminary analyses of the association between program length and ES_{RM} compared linear, quadratic, logarithmic, power, and growth models as mathematical representations of this relationship. The logarithmic model given as Equation 1 provides the best prediction¹ of ES_{RM} (t is program length, in weeks):

$$ES_{RM} = 1.10 + .42 * \ln(t) \quad (1)$$

The graph of this equation is given in Figure 1. The correlation of ES_{RM} with program length was small ($r = .18$), but statistically significant ($\chi^2 = 27.63$, 1 *df*, $p < .001$). It is important to note that the linear form of the model could be misleading. The intercept (1.10) might be mistakenly interpreted as indicating that ES_{RM} is predicted to be greater than 0 prior to training (at $t = 0$). If this equation expressed a simple linear regression of ES_{RM} on weeks of training, then this would be the usual interpretation of the equation intercept. This interpretation is misleading because the equation takes as input the natural logarithm of time (number of weeks). If we solve the equation for an ES_{RM} of 0, the estimated time to produce an effect of this size is .52 days. This estimate would likely correspond to, at most, one training session in a typical program.

¹ As measured by the Akaike Information Criterion (AIC). The AIC is a widely used tool for model selection that incorporates the fit of a model to the data (specifically, the probability of the data given the model), and the complexity of the model, which is measured by the number of parameters in the model and is incorporated to penalize overfitting.

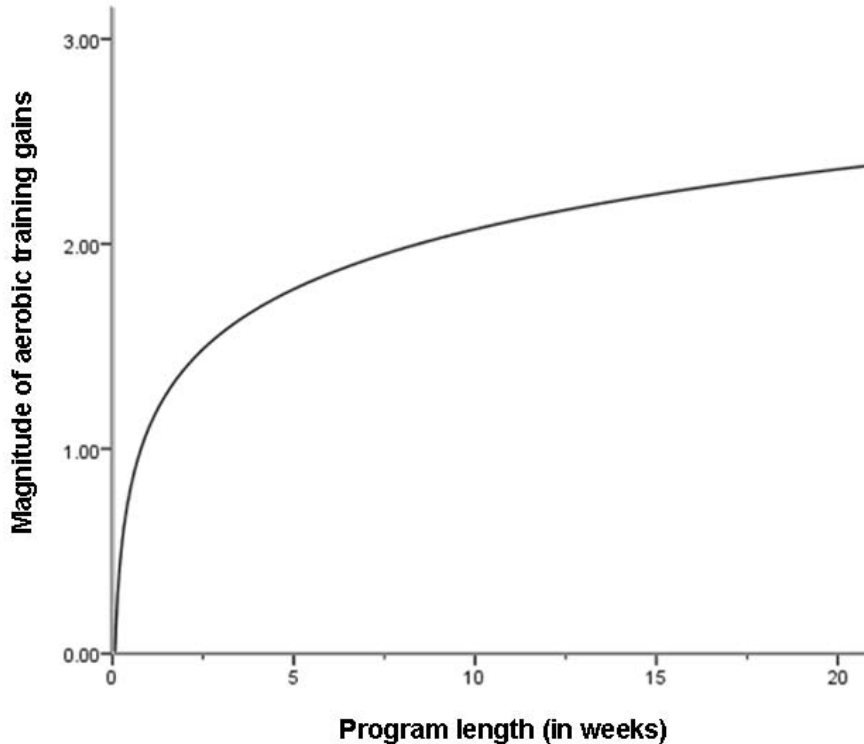


Figure 1. Gain in aerobic training response, ES_{RM} , as a function of program length.

Initial Moderator Analyses

The overall mean ES_{RM} due to aerobic training was 2.03; however, the analysis revealed significant heterogeneity among the effect sizes, so this single value provides a poor summary description of the data. A more accurate description will rely on an investigation of potential moderating variables that can help explain the variation between studies. To this end, moderator analyses were conducted treating ES_{RM} as the dependent variable.

A summary of the initial moderator analyses is provided in Table 3, which includes the χ^2 and TLI values for each of the moderator variables. Table 4 (p. 16) summarizes the program design moderator effects in terms of best practices. Table 4 provides the average effect sizes for different options for each program design facet, and indicates the equivalence sets based on those averages. The program design facets of intensity, frequency, and duration were significant moderators, as expected from previous reviews, whereas program type (intermittent or continuous) was not. Exercise type was a statistically significant moderator, but the TLI value indicated that the differences were trivial. This moderator will be excluded from subsequent analyses. Of the participant characteristics, gender group and initial fitness level were statistically significant, with the latter being a particular strong moderator of ES_{RM} . However, the influence of gender group is driven almost entirely by study samples consisting of both men and women; samples composed of either just men or just women did not differ significantly. In what follows, the most relevant moderator variables from a program design perspective are discussed in greater detail. In addition, more focused moderator analyses will be conducted that take into account the prevalent influence of initial fitness level as a key demographic moderator variable.

Table 3
Summary of Overall Moderator Analyses

Moderator	χ^2	<i>df</i>	Sig	TLI
Age	4.50	3	0.213	<.001
Gender group	49.95	2	<0.001	.065
Men vs. women	2.82	1	0.093	<.001
Initial fitness level	104.84	6	<0.001	.128
Intensity	44.65	2	<0.001	.066
Frequency	45.84	4	<0.001	.050
Duration	40.04	4	<0.001	.045
Program length	67.80	6	<0.001	.080
Exercise type	7.96	3	0.047	<.001
Type of program	0.68	2	0.713	<.001

Intensity. The intensity of a training program was a statistically significant moderator ($\chi^2 = 44.65$, 2 *df*, $p < .0001$, TLI = .07). As intensity increased, ES_{RM} increased monotonically, and (almost) linearly. Training intensities that varied from *very hard to maximal* produced the largest overall gain (as shown in Figure 2); a qualitative pattern that is consistent with previous reviews. Post hoc comparisons revealed that the effect size for the most intense training differed significantly from hard intensity ($\chi^2 = 22.00$, 1 *df*, $p < .0001$), implicating *very hard to maximal* intensity as a best practice.

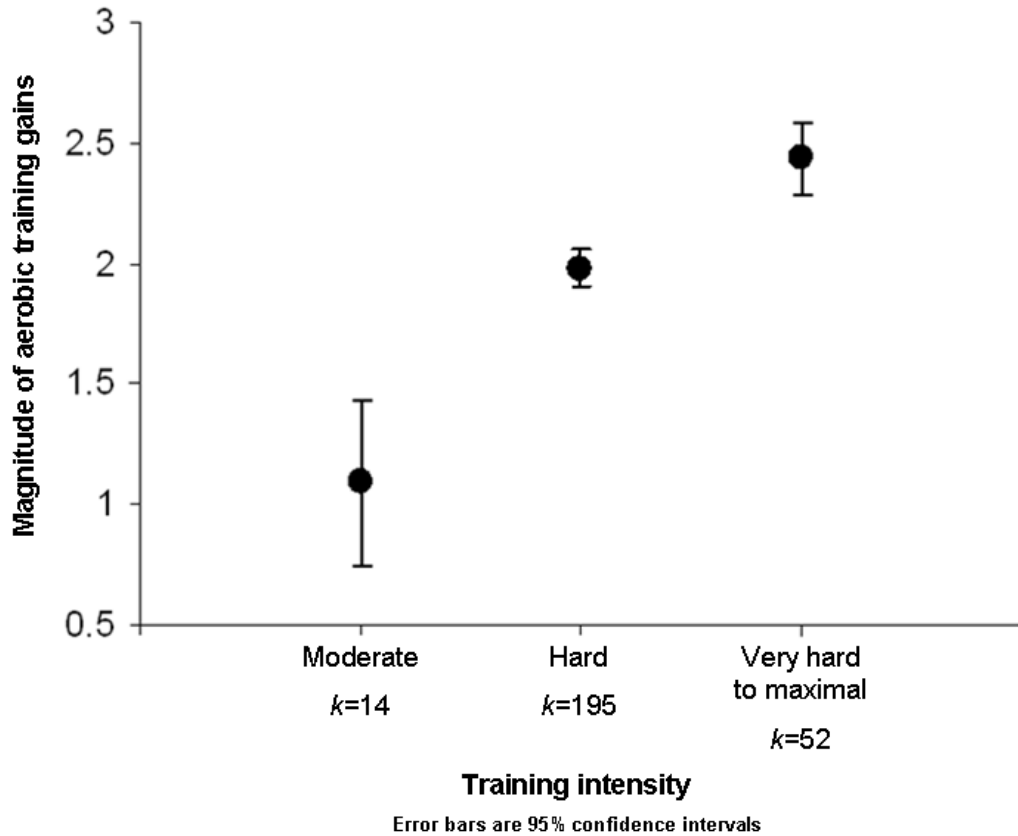


Figure 2. Gain in aerobic training response (as indicated by the ES_{RM} depicted on the y-axis) as a function of the intensity of the training program. Each level of the independent variable is labeled by the number of effect sizes (k) associated with that level.

Frequency. The number of sessions per week was a statistically significant moderator ($\chi^2 = 45.84$, 4 df , $p < .0001$, $TLI = .05$). This preliminary analysis included all data points, but the average ES_{RM} for 1-2 days per week exceeded the gains of any other number of days per week. This counterintuitive result is illustrated in Figure 3, and is driven by two very large ES_{RM} (> 4). A boxplot examination of these data suggests that they are mild outliers (that is, just outside the interquartile range), so a second analysis was conducted with these two ES_{RM} removed. With these data points removed, frequency was still a statistically significant moderator ($\chi^2 = 32.26$, 4 df , $p < .0001$, $TLI = .03$), with the largest ES_{RM} associated with 4 sessions per week. The qualitative pattern of results, shown in Figure 3, is consistent with previous reviews (i.e., Wenger & Bell, 1986). Post hoc comparisons of the trimmed data revealed that 4 sessions per week differed significantly from 3 sessions per week, but not from 4, 5, or 1-2 sessions per week. Care should be taken, however, in evaluating the relative benefits of training 1 to 2 times per week versus 4 or 5 times, as only 18 ES_{RM} were included in the 1-2 sessions group, which may be too few to reach any strong statistical conclusion.

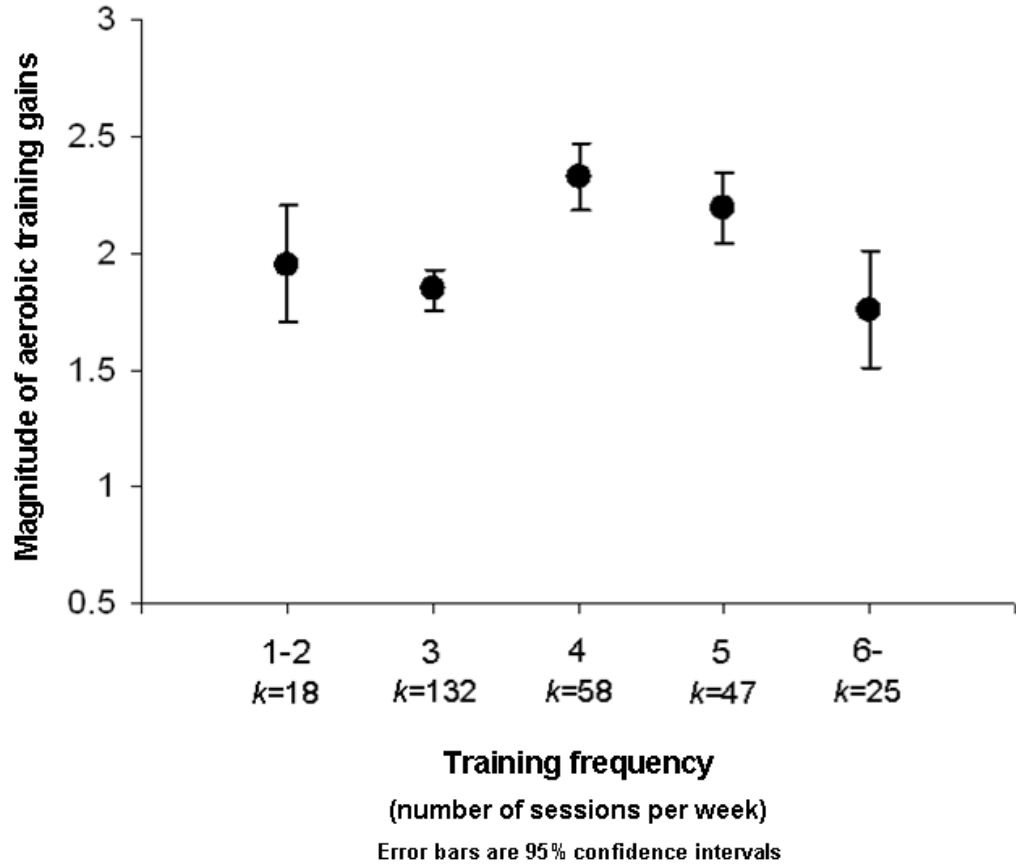


Figure 3. Gain in aerobic training response (as indicated by the ES_{RM} depicted on the y-axis) as a function of frequency of training (in sessions per week).

Duration. The duration of a training session was a statistically significant moderator ($\chi^2 = 40.04, 4 df, p < .001, TLI = .05$). Training durations exceeding an hour produced the largest training response ES_{RM}, but this value differed significantly only from training durations of less than 15 minutes ($\chi^2 = 27.38, 1 df, p < .001$). The relationship between aerobic training gains and session duration is illustrated in Figure 4.

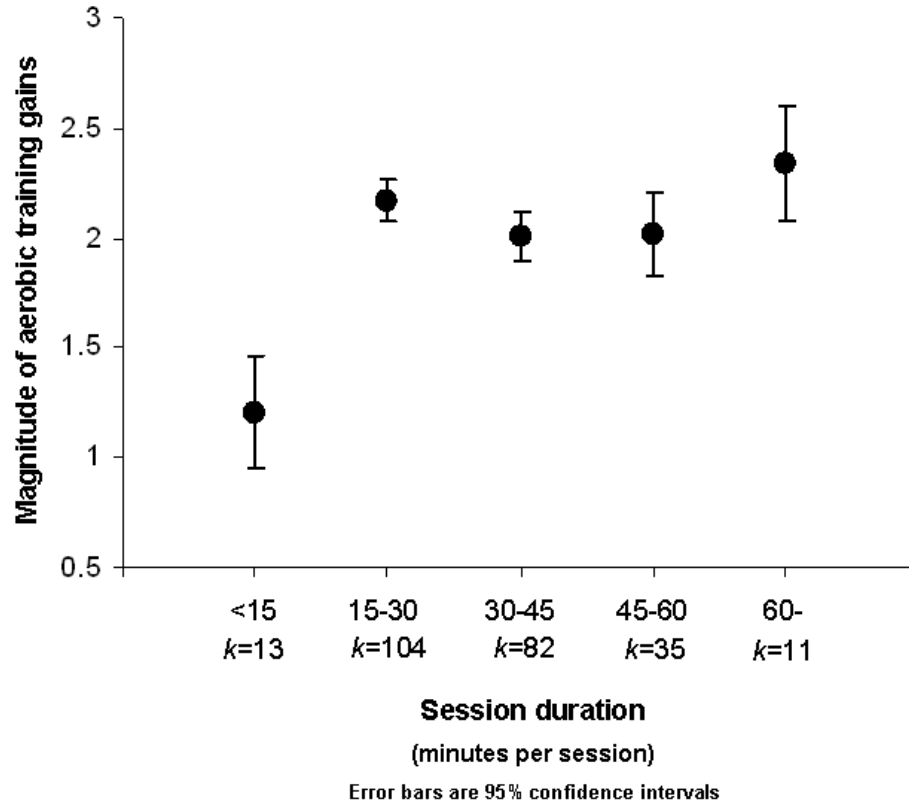


Figure 4. Gain in aerobic training response as a function of duration. Each level of the independent variable is labeled by the number of effect sizes (*k*) associated with that level.

Initial fitness level. The initial fitness level of study participants was a key moderator variable ($\chi^2 = 104.84, 6 df, p < .0001, TLI = .13$). As shown in Figure 5, though gains in aerobic capacity were achieved across all levels of initial fitness, the largest gains were obtained by study participants of relatively modest initial fitness. The relationship between initial fitness level and aerobic fitness gains is not unexpected; indeed, the same relationship has been found for resistance training (Vickers, Hervig, & Barnard, unpublished report).

ES_{RM} generally increased with program length for both unfit ($\chi^2 = 53.49, 6 df, p < .0001, TLI = .07$) and fit individuals ($\chi^2 = 15.29, 6 df, p = .02, TLI = .11$). The logarithmic models for unfit and fit individuals are given below as Equations 2 and 3, respectively:

$$ES_{RM} = 1.91 + .22 * \ln(t) \tag{2}$$

$$ES_{RM} = .70 + .49 * \ln(t) \tag{3}$$

Moderator Analyses Adjusted for Initial Fitness Level

The bivariate moderator analyses revealed expected effects of key training program facets. But the strong influence of initial fitness on ES_{RM} suggests that a more focused moderator analysis would hold initial fitness level constant. To this end, participants were divided into two groups, and analyses were carried out for each group separately. The first group (*N* = 202) included all

participants classified as being of very poor to average fitness (“unfit”), and the second ($N = 92$) included all those classified as being of good or better fitness (“fit”) using the classification standards found in Appendix C. To preview, with the exception of program length, significant moderator effects held only for unfit individuals; accordingly, a detailed summary of the results for unfit individuals is provided in Table 5 (p. 19).

Intensity. Program intensity was a statistically significant moderator for unfit individuals ($\chi^2 = 47.91, 2 df, p < .0001, TLI = .08$), but not for fit individuals ($\chi^2 = 2.74, 2 df, p = .25$). As in the overall analysis, *very hard to maximal* training intensities produced the largest overall gain. Post hoc comparisons revealed that the effect size for the greatest intensity differed significantly from hard intensity ($\chi^2 = 23.16, 1 df, p < .0001$), implicating *very hard to maximal* intensity as a best practice for unfit individuals.

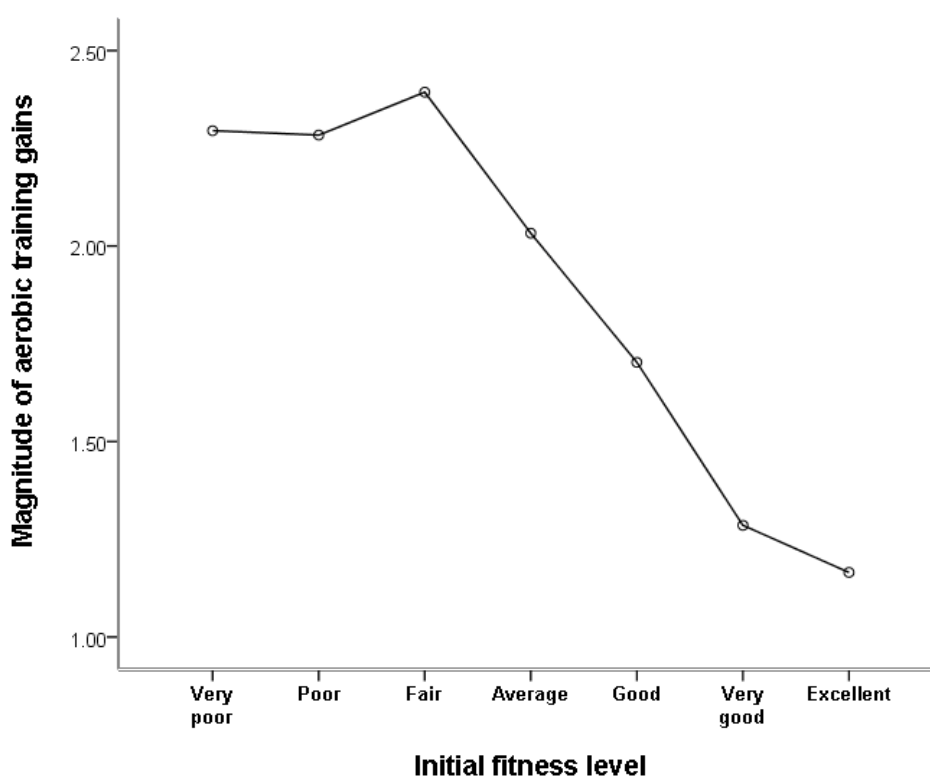


Figure 5. Gain in aerobic training response, ES_{RM} , as a function of initial fitness level.

Frequency. The number of sessions per week was a statistically significant moderator for unfit individuals ($\chi^2 = 34.39, 4 df, p < .0001, TLI = .10$), but not for fit individuals ($\chi^2 = 5.09, 4 df, p = .28$). In contrast to the analysis of the overall results, the largest gains occurred for 6 sessions per week, but post hoc comparisons revealed that this value differed significantly only from 3 sessions per week ($\chi^2 = 7.76, 1 df, p < .01$).

Duration. The duration of a training session was a statistically significant moderator for unfit individuals ($\chi^2 = 61.12, 4 df, p < .001, TLI = .10$), but not for fit individuals ($\chi^2 = 2.93, 4 df, p = .57$). As in the overall analysis, training durations exceeding an hour produced the largest training

response ES_{RM} . This value differed significantly from the second largest ES_{RM} , associated with 15 to 30 minutes ($\chi^2 = 15.10$, 1 *df*, $p < .001$), implicating durations longer than an hour as a best practice. However, care should be taken in the interpretation of this result. Only 6 ES_{RM} were included in durations longer than an hour, and 2 of those 6 data points were exceptionally large (greater than 4) and outside the interquartile range, suggesting that those data points are likely outliers. Removing those data points reduces the ES_{RM} from 3.10 to 1.49, highlighting their influence on the overall analysis. With those influential data points removed, duration was still a statistically significant moderator for unfit individuals ($\chi^2 = 46.73$, 4 *df*, $p < .001$, TLI = .08), but with session durations between 15 and 60 minutes showing the largest gains, illustrated in Figure 6. Post hoc comparisons revealed that the average ES_{RM} for 15 to 30 minutes differed significantly from durations exceeding an hour ($\chi^2 = 6.27$, 1 *df*, $p = .01$), suggesting that 15 to 60 minute durations are an equivalence class.

Table 4
Design Facet Moderator Effects

Moderator	Level	ES_{RM}	k ^a	Equivalence set ^b
Intensity	Moderate	1.09	14	
	Hard	1.98	195	
	Very hard to maximal	2.43	42	
Frequency (number of sessions per week)	1-2	1.95	18	
	3	1.85	132	
	4	2.33	58	(4, 5, 1-2}
	5	2.20	47	
	> 6	1.76	25	
Duration (minutes per session)	< 15	1.21	13	
	15-30	2.17	104	{>61, 16-30, 46-60, 31-45}
	30-45	2.01	82	
	45-60	2.02	35	
	> 60	2.33	11	

^a“k” is the number of samples that provided averages for analysis. ^bThe equivalence sets include all design options that were not significantly different from the option with the highest ES_{RM} . The design options are listed from largest to smallest ES_{RM} in the set.

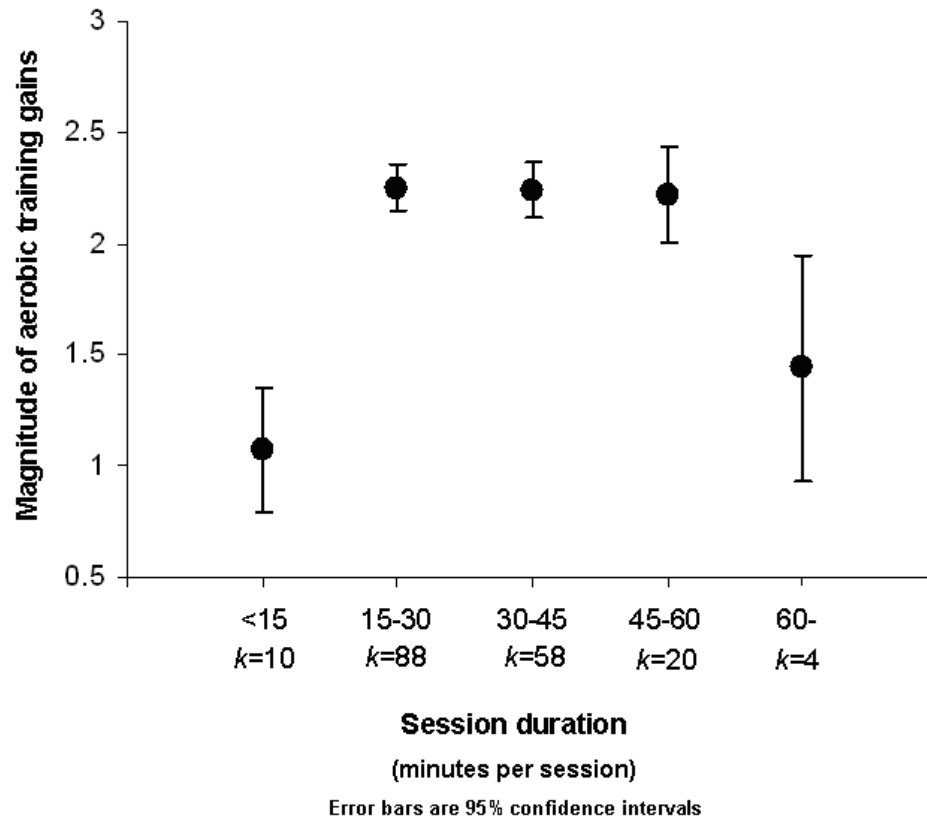


Figure 6. Gain in aerobic training response, ES_{RM} , for unfit individuals as a function of session duration (after removing influential data points from durations exceeding an hour). Each level of the independent variable is labeled by the number of effect sizes (k) associated with that level.

Discussion

Changes in aerobic capacity depend on several factors, some related to the characteristics of individuals, others to the characteristics of the training programs. Within the latter set, this review confirms previous research that has demonstrated the importance of intensity, frequency, duration, and program length as factors that contribute to aerobic fitness. However, with the exception of training intensity, this review has not identified best practices. Program design facets were statistically significant moderators of ES_{RM} , but post hoc analyses did not single out one option as significantly better than all others. Failing to identify best practices is not unique to this review. Program design facets often are statistically significant moderators of the training response, but post hoc analyses fail to identify any single option as significantly better than all other options. Given this general trend, the current findings could not be dismissed as resulting from the inclusion criteria or analysis procedures that have been employed in the current review. Nor should the current findings indicate that perhaps there is something unique about aerobic training that precludes the identification of best practices: a recent meta-analysis on resistance training (Vickers, et al., unpublished report) also failed to identify best practices using the same criterion.

The statistical methods adopted in this review should have sharpened the contrasts between design options. Specifically, repeated measures analyses are expected to produce larger effect sizes, since repeated measures designs give rise to smaller sampling variability. The smaller sampling variability amplifies differences between the average ES_{RM} values for different design options in post hoc analyses, so the current procedure should have increased the likelihood of finding a best practice.

The failure to identify best practices does not mean that such practices do not exist. Every analysis produced one option that had a larger ES_{RM} than all other options for that facet. The problem was that the differences between the most promising option and other choices were not large enough to be statistically significant. Although the comparisons have not been reported in detail here, many post hoc comparisons produced very small χ^2 values despite moderately large sample sizes. The implication is that the available evidence would have to be multiplied many times to make the contrasts between the design options statistically significant. If the required data were available, the conclusion still might be that the differences were too small to be important. It is debatable whether the extensive additional research needed to clearly define best practices would really have much impact on program design choices.

A low probability of identifying best practices at any time in the near future does not mean that aerobic training research fails to offer any advice on training program design. Aerobic training research helps to single out some design options as less effective than others. While the typical equivalence set included more than one option, it is also true that it seldom contained all possible options. Given the available evidence, trends in the data that are corroborated across different reviews may suggest sound practical guidelines, subject to constraints that a program coordinator might face (e.g., the cost, in terms of dollars or time, to implement one facet instead of another). As a guideline for future studies, it may be more productive to conduct research to rule out some options—focusing on what is *reasonable*, given what we do know, than on what is *best*, absent what is almost impossible to know.

Finally, comparing the results of this meta-analysis to a recent review of the resistance training literature (Vickers, et al., unpublished report) could potentially yield general training principles that can help to inform reasonable expectations for any physical training program. For example, in this meta-analysis and in Vickers, et al., unfit individuals showed significantly greater improvement than fit individuals. Also, the rate of improvement for both aerobic and resistance training followed a similar growth pattern, one best described mathematically as a logarithmic function.

Specifically, after comparing the growth patterns associated with aerobic and resistance training, some striking similarities emerge. First, the direction and strength of the relationship between program length and ES_{RM} are nearly identical across training types ($r = .18$ and $r = .21$ for aerobic and resistance training, respectively). Second, the best statistical model relating program length and ES_{RM} shares an identical structural form and similar set of estimated parameters across training types. In particular, these equations are:

$$ES_{RM} = 1.10 + .42 * \ln(t) \quad (\text{Aerobic training; } 4)$$

$$ES_{RM} = 0.41 + .55 * \ln(t) \quad (\text{Resistance training; } 5)$$

Table 5
Design Facet Moderator Effects for Unfit Individuals

Moderator	Level	ES _{RM}	k	Equivalence set
Intensity	Moderate	1.13	11	
	Hard	2.11	142	
	Very hard to maximal	2.63	35	
Frequency (sessions per week)	1-2	2.22	12	(>6, 4, 5, 1-2}
	3	1.94	97	
	4	2.53	43	
	5	2.24	38	
	> 6	2.60	9	
			$\chi^2 = 47.91, 2 df$ $p < .0001, TLI = .08$	
Duration (in minutes)	< 15	1.13	10	(16-30,31-45, 46-60}
	16-30	2.26	88	
	31-45	2.24	58	
	46-60	2.21	20	
	> 61	3.10/1.49 ^a	6/4 ^a	
			$\chi^2 = 34.39 df,$ $p < .0001, TLI = .05$	
Program length (in weeks)	1-2	1.07	1	{9-10, 3-4, >14}
	3-4	2.61	6	
	5-6	1.97	21	
	7-8	1.99	45	
	9-10	2.68	45	
	11-13	1.96	41	
	> 14	2.41	37	
			$\chi^2 = 61.12, 4 df,$ $p < .0001, TLI = .10$ $\chi^2 = 46.73, 4 df,$ $p < .0001, TLI = .08^b$	
		$\chi^2 = 53.49, 6 df,$ $p < .0001, TLI = .07$		

^aLatter values are the ES_{RM} and sample size for the group after removing influential data points.

^bSummary statistics for the analysis after removing influential data points from the “>61” group.

The equations share similar rates of change in ES_{RM} , implying that aerobic and resistance gains accrue at roughly the same rate. The equations differ primarily with respect to the intercept, an additive constant. To determine what this difference implies, note again that the intercept here does not have the usual interpretation found in linear regression (i.e., the value of the dependent variable when the independent variable is set to zero). Since the predictor is transformed logarithmically, one way to interpret the intercept is to solve the equation for $ES_{RM} = 0$. The solution is an estimate of the number of sessions expected to produce no training effect. For resistance training, the number of sessions producing no effect corresponds to 1 or 2 in a typical program; for aerobic training, the number of sessions corresponds to one at most. In other words, aerobic training could be expected to return measurable gains a bit earlier in a typical program than strength training, but over time the overall rate of return on training would differ only slightly between the two program types.

Practical Recommendations?

The primary aim of this review has been to determine whether best practices exist for different program design facets. It is important to note that the criterion for what counts as a best practice was a stringent one, and that in many cases there may simply have been too little power to detect a potentially important difference. While the review could end here, treating the null results as a call for further research, it is important not to lose sight of what reviews such as these are intended to achieve; namely, informed suggestions for guidelines given the available evidence.

But with no clear evidence that best practices exist for most design facets, what procedures are available to translate effect size estimates into practical guidelines? One way is to translate the results into a more intuitively meaningful measure that could provide a secure basis (if not ideal) for program design choices. This review has focused on average individual change for a given study sample, and employed ES_{RM} as the measure of improvement in aerobic fitness. While this measure is appropriate for meta-analyses that focus on individual change (and employ a pretest-posttest design), the results can sometimes be difficult to interpret.

For example, how much better is a training facet that yields an ES_{RM} of 2.46 compared to one that yields 2.25? It may be the case that the difference is statistically significant, which would implicate the former facet as a better practice than the latter, but how much should we read into such a difference?²

One way to answer this question is to translate effect sizes into an estimated percentage of the trained population that would be expected to improve their cardiorespiratory fitness. For example, an ES_{RM} of .65 implies that the change would be positive for 74% of program participants (Morris & DeShon, 2002).³ For 2.46 and 2.25, the estimates are 99.3% and 98.8%, suggesting that we should not read too much into the difference between the effect sizes. The main drawback to this

² A similar problem has been addressed earlier in the review, which motivated the use of the TLI to estimate the importance of a statistically significant finding (see Appendix B for details). However, the goal of this section is to connect the results of the meta-analysis to practical guidelines, and the TLI does not admit a natural interpretation that would address this problem.

³ Assuming normally distributed data. See Morris and DeShon (2002) for a more detailed discussion of this assumption.

approach is that it only estimates the percentage of participants expected to improve, but not by how much, relative to those who had not trained at all. Since *some* improvement would be expected in response to any training, this strategy may be uninformative. Indeed, for nearly all training facets presented in Tables 4 and 5, estimated percentages ranged from 96%-99%; the only exceptions were moderate intensity training (86% expected to improve), and training for less than 15 minutes per session (89% expected to improve).

The main limitation of the previous method is due to the fact that it relies on an effect size that is based on the variability of change scores, which focuses on individual change and not on the relationship between a trained group and an untrained group. Thus, another way to answer the question is to shift the research focus to an analysis based on score variability within the separate groups. The latter analysis enables us to make statements about the average performance of one group relative to the other. Assuming that the populations are normally distributed with equal variance, we can then translate effect size estimates into percentile rankings (e.g., the average performance after training at a hard intensity was greater than 81% of the no-training population).

To shift from a focus on individual change scores to a comparison between groups, the ES_{RM} was converted to the effect size for independent groups (ES_{IG}) (Morris & DeShon, 2002). For the ES_{RM} of 2.46 and 2.25 the corresponding ES_{IG} are 1.10 and 1.01 (the conversion formula—Equation A3—and its meaning are provided in Appendix A). An ES_{IG} of 1.10 means that the average performance of the trained group was better than 86% of the untrained population; an ES_{IG} of 1.01 means that the average performance of the trained group was better than 84% of the untrained population. As these numbers suggest, perhaps we should not read too much into the difference between the groups. The estimated percentages for the main program design facets are provided in Table 6.

These percentages are intended to complement the more stringent statistical definition of a “best practice.” However, despite improvement in the interpretability of the results, open questions remain. While it may be apparent that it is worthwhile to design a training program with at least a hard intensity level (given the 14% relative increase over moderate intensity training), it is less clear in other cases. Would the choice for a *very hard to maximal* training intensity be warranted, given the 5% relative increase, but also greater risk for injury? This and related questions are beyond the scope of this review—they entail understanding the physical fitness demands expected for particular occupations, and the larger economic and environment contexts in which training will occur. Ideally, these percentages would be judged relative to occupational physical fitness standards. Future research should be directed towards understanding how to relate standardized measures of aerobic gains to the physical demands of specific occupations.

Table 6
Design Facet Moderator Effects for All and Unfit Individuals

Moderator	Level	All individuals ^a	Unfit individuals ^b
Intensity	Moderate	69%	69%
	Hard	81%	83%
	Very hard to maximal	86%	88%
Frequency (number of sessions per week)	1-2	81%	84%
	3	80%	81%
	4	85%	87%
	5	84%	84%
	> 6	78%	88%
Duration (minutes per session)	< 15	71%	69%
	16-30	83%	84%
	31-45	82%	84%
	46-60	82%	84%
	> 61	85%	75%
Program length (in weeks)	1-2	67%	68%
	3-4	76%	88%
	5-6	80%	81%
	7-8	80%	81%
	9-10	86%	88%
	11-13	80%	81%
	> 14	85%	86%

^a Percentile ranking that the average trained individual would have in the untrained population (data are from all participants—no distinction in fitness made). ^b Percentile ranking that the average trained individual would have in the untrained population (data are only from unfit individuals).

As a further guide toward practical recommendations, it should be noted that the conclusions of this analysis are generally consistent with the recommendations of the American College of Sports Medicine (ACSM). For healthy adults under the age of 65, the ACSM recommends moderate to intense training, 20 to 30 minutes a day, 3 to 5 days a week. In addition, the ACSM emphasizes the point that physical activity exceeding the basic recommendations provides even greater health

benefits (a finding corroborated by this meta-analysis). Comparing the ACSM's recommendations to the results of this meta-analysis highlights broad commonalities, with increasing levels of intensity yielding larger results, the greatest gains from weekly frequency occurring at 4 sessions per week, and session durations between 20 to 30 minutes returning the largest absolute gain.⁴ But an important advantage of this meta-analysis over the ACSM and previous meta-analyses is that it provides quantitative estimates of the relative expected gains from several design facets. While design options could not be distinguished in most cases on the basis of best practices, these estimates will play an important role in developing statistical models for predicting expected aerobic gain, given a set of chosen design features.

⁴ There are essentially no differences in fitness gains among the duration groups of 15-30 minutes, 30-45, and 45-60 minutes.

Appendix A

Computing Effect Sizes for Repeated Measures

Meta-analysis provides estimates of the average ES, and the variation of individual ES estimates about that average. The homogeneity tests for variation about the average are especially important in the present context. If the ESs for different training programs display greater-than-chance variation, it is reasonable to search for moderator variables that can explain the observed heterogeneity. In the present review, program design facets and demographic variables were of interest as potential moderator variables.

Studies must be assigned appropriate weights to compute the average ES and test for variation about the average. The weights are based on the precision of the individual ES estimates. All studies reviewed here employed pretest-posttest research designs. In such cases, the correlation of pretest scores with posttest scores affects the sampling variance that is the index of precision for the ES estimate. Therefore, the pretest-posttest correlation must be known to derive sampling variance estimates that are suitable for determining ES weights. The correlation must be known whether the analyses employ standardized mean change scores or difference scores (Morris, 2000). For change scores, the proper estimate of sample variance is:

$$\sigma_{\text{Diff}}^2 = \sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2. \quad (\text{A1})$$

In this equation, the subscripted “Diff” indicates that the variable of interest is a difference score. The pretest-posttest correlation, r , is expected to be positive and moderate to large. As a consequence, the last term of Equation A1 will be moderate to large relative to the first two terms. It follows that simply pooling the pretest and posttest variances, as would be the case if the pretest-posttest correlation was ignored, will result in overestimation of the true sampling variance. If the variance is overestimated, the z -scores associated with the deviation of specific ES values from the average ES will be smaller than they would be if the correct variance were used. The overall test for homogeneity of ESs, Cochran’s Q , is the sum of the squared z -scores. Thus, overestimating sampling variance will lead to underestimating Q . This bias in the Q -test values could lead erroneously to the conclusion that a given moderator is unimportant. The tests for moderators were central to this review, so accurate variance estimates were essential.

The correct variance estimates could be estimated easily if studies routinely reported the pre training/post training correlations for test scores. Unfortunately, this information is seldom reported. The required information could be extracted from the t -test or F -test if either statistic was reported separately for each condition in the study. Once again, aerobic training studies seldom provide this information.

After developing pretest-posttest correlation estimates, the analysis followed guidelines provided by Morris and DeShon (2002). First, the variance for individual observations was computed by applying Equation A1 above. Second, the standard deviation of the differences (SD_{diff}) was computed by taking the square root of the variance. This standard deviation was used to compute the initial ES_{RM} (Equation A2). A separate ES was computed for each record in the data file. A record consisted of the results for a single aerobic training program administered to a particular sample of subjects.

The use of an average pretest-posttest correlation will obviously be inaccurate in many cases. However, these correlations clearly have been positive and substantial when estimates have been available. Ignoring this strong trend would lead to very conservative tests for moderator effects. The uncertainty introduced by the use of average values was preferable to having results that certainly were too conservative.

The estimated pretest-posttest correlation values were combined with the sample standard deviations to compute the variance of the difference scores as shown in Equation A1. SD_{diff} was the square root of this variance. The ES for repeated measures was

$$ES_{RM} = (\text{Mean}_{post} - \text{Mean}_{pre}) / SD_{diff} . \quad (A2)$$

Equation A2 depends on the estimated variability associated with the change scores (i.e., the SD_{diff}). The ES_{RM} is appropriate in cases in which one is interested in the average improvement due to a training program (measured in standard deviation units) above zero (Morris & DeShon, 2002). However, if one is interested in the average performance due to training relative to the average performance without training, then the variability associated with the separate groups is used as the basis for the effect size calculation. The formula for converting an ES for repeated measures to an ES for independent groups was

$$ES_{IG} = ES_{RM} \sqrt{2(1-\rho)} . \quad (A3)$$

Weighting ES_{RM} Estimates

Individual ES estimates must be weighted to obtain the most precise aggregated ES_{RM} estimate and to test for heterogeneity in the individual estimates. The appropriate weights are the inverse of the variance. The variance of an individual ES_{RM} estimate can be computed by applying the equation for the single-group pretest-posttest change score variance formula in Table 2 of Morris and DeShon (2002, p. 117).

$$\text{Variance} = \left(\frac{1}{n} \right) \left(\frac{n-1}{n-3} \right) \left(1 + n\delta_{RM}^2 \right) - \left(\frac{\delta_{RM}^2}{c^2} \right) . \quad (A4)$$

In this equation, n is sample size and δ_{RM} is the population value for ES_{RM} . The equation includes a bias correction, c , to obtain accurate variance estimates. This correction factor was obtained by applying the approximation developed by Hedges (1982), and given as Equation 23 in Morris and DeShon (2002, p. 117)

$$c = 1 - \frac{3}{(4 * df) - 1} . \quad (A5)$$

The variance computations required one additional input, δ_{RM} . Ideally, this parameter would be set equal to the unknown population ES. An estimate of this population parameter, d_{RM} , was used because the population value can only be estimated once after the variance is already known.

Given this circularity, the recommended solution is to compute the unweighted ES and use that value for computing the variance for ES_{RM} (Hedges, 1982; Morris & DeShon, 2002). The present analyses employed this approach.

The variance of ES_{RM} was computed by applying Equation A4 after estimating δ_{RM} . The derivation of that equation can be found in the appendix to Morris and DeShon (2002), or in Gibbons, Hedeker, and Davis (1993).

Appendix B Tucker-Lewis Index

The TLI (Tucker & Lewis, 1973) was introduced to guard against what may be a wide-spread problem in meta-analysis. Moderator analyses begin with a significance test. If the null hypothesis is rejected, the moderator variable is accepted as a meaningful influence on ES even if the differences between groups are quite small.

Relying on significance tests to identify important results is a risky proposition in any statistical analysis. Statistical significance is the product of sample size and ES (Rosenthal & Rosnow, 1984). In meta-regression, ES might be labeled “meta-ES” because it reflects the differences in the primary ES across moderator groups. A significant meta-ES could indicate a substantial between-groups difference, but it does not rule out the possibility that small between-groups difference have been amplified by a large sample size. Although it follows that the meta-ES must be separated from sample size to properly interpret findings, this principle is not routinely applied to meta-analysis even though logic says it should be.

The TLI was adapted to provide a meta-ES metric. The TLI, which is the proportion of greater-than-chance variation in ES , can be computed from the χ^2 values from a moderator analysis. The variation in ES_{RM} determines the χ^2 values. The TLI equation is

$$TLI = \frac{(\chi_{Null}^2 / df_{Null}) - (\chi_{Model}^2 / df_{Model})}{(\chi_{Null}^2 / df_{Null}) - 1} \quad (B1)$$

The expected value of χ^2/df ratio is 1, so the denominator of Equation B1 is the proportion of the observed variation in ES_{RM} that is greater than expected by chance. The numerator is the variation in ES_{RM} accounted for by the model, i.e., total ES_{RM} variation minus the residual ES_{RM} variation after fitting the model. The TLI is a reasonable index of the meta-regression ES, and maintains a connection between effect size and the probability that a moderator will be statistically significant.

The TLI is not an exact parallel to the usual effect size indicators such as the proportion of variance explained in an ANOVA. One reason is that the TLI is analogous to Hays' (1963) ω^2 rather than the usual ε^2 . The difference between the two is that the variance that would be expected by chance is subtracted from the variance explained when computing ω^2 , but not when computing ε^2 . This difference is the reason that i will be less than zero when $\chi_{Model}^2 / df_{Model} > \chi_{Null}^2 / df_{Null}$ because the numerator will be a negative number. This situation arises when the reduction in the χ^2 produced by a model is small relative to the number of parameters in the model. For this reason, the reported TLI is the value derived from Equation B1 or .00, whichever is larger.

The interpretation of the TLI employed Cohen's (1988) general criteria for ES evaluations. Cohen's criteria classify ESs on the basis of the proportion of observed variation explained by a predictor. In this case, TLI is the proportion of non-random variation in ES_{RM} , so Cohen's (1988) ES classification rule is a suitable index for characterizing the strength of association of moderator variables with ES_{RM} : small meta-ES, $.01 \leq TLI < .10$; moderate meta-ES, $.10 \leq TLI < .25$; large meta-ES, $TLI \geq .25$.

Appendix C
Coding schemes for intensity and initial fitness level

The classification schemes for levels of intensity (for distinct physiological and behavioral measurements), and initial fitness level are presented below. For intensity, the columns represent percentage of \dot{V}_{O_2} reserve (%VO2R) and percentage of heart rate reserve (%HRR); percentage of maximum heart rate (%HRmax); percentage of $\dot{V}_{O_{2max}}$ (%VO2max); and ratings of perceived exertion (RPE). This classification scheme is based on Heyward (2006).

The initial fitness codes are based on pre training $\dot{V}_{O_{2max}}$ values, and vary depending on age and gender (Source: http://preventdisease.com/news/articles/vo2_max_how_fit_athlete.shtml).

Table C1
Intensity coding scheme

Intensity	% VO2R or %HRR	%HR max	% VO2max	RPE
Very light	<20	<35	<1	<10
Light	20-39	35-54	2-27	10-11
Moderate	40-59	55-69	28-50	12-13
Hard	60-84	70-89	51-81	14-16
Very hard	85+	90+	82+	17-19
Maximal	100	100	98	20

Table C2
Initial Fitness Level coding scheme

Age (years)	Very poor		Poor		Fair		Average	
	Men	Women	Men	Women	Men	Women	Men	Women
20-24	<32	<27	32-37	27-31	38-43	32-36	44-50	37-41
25-29	<31	<26	31-35	26-30	36-42	31-35	43-48	36-40
30-34	<29	<25	29-34	25-29	35-40	30-33	41-45	34-37
35-39	<28	<24	28-32	24-27	33-38	28-31	39-43	32-35
40-44	<26	<22	26-31	22-25	32-35	26-29	36-41	30-33
45-49	<25	<21	25-29	21-23	30-34	24-27	35-39	28-31
50-54	<24	<19	24-27	19-22	28-32	23-25	33-36	26-29

Note. Ranges are pre-training $\dot{V}_{O_{2max}}$ values.

Table C2
Initial Fitness Level coding scheme

Age (years)	Good		Very good		Excellent	
	Men	Women	Men	Women	Men	Women
20-24	51-56	42-46	57-62	47-51	>62	>51
25-29	49-53	41-44	54-59	45-49	>59	>49
30-34	46-51	38-42	52-56	43-46	>56	>46
35-39	44-48	36-40	49-54	41-44	>54	>44
40-44	42-46	34-37	47-51	38-41	>51	>41
45-49	40-43	32-35	44-48	36-38	>48	>38
50-54	37-41	30-32	42-46	33-36	>46	>36

Acknowledgments

We thank Amanda Barnard for her help in the literature search, and David P. Hampton Jr for his assistance in gathering research reports and in data file construction.

References

- * Article provided data for the meta-analysis.
- *Acevedo, E. O., & Goldfarb, A. H. (1989). Increased training intensity effects on plasma lactate, ventilatory threshold, and endurance. *Med Sci Sports Exerc*, 21(5), 563-568.
- *Adams, T. D., Yanowitz, F. G., Fisher, A. G., Ridges, J. D., Lovell, K., & Pryor, T. A. (1981). Noninvasive evaluation of exercise training in college-age men. *Circulation*, 64(5), 958-965.
- *Allemeier, C. A., Fry, A. C., Johnson, P., Hikida, R. S., Hagerman, F. C., & Staron, R. S. (1994). Effects of sprint cycle training on human skeletal muscle. *J Appl Physiol*, 77(5), 2385-2390.
- *Allen, D., Freund, B. J., & Wilmore, J. H. (1986). Interaction of test protocol and horizontal run training on maximal oxygen uptake. *Med Sci Sports Exerc*, 18(5), 581-587.
- *Andersen, P., & Henriksson, J. (1977). Capillary supply of the quadriceps femoris muscle of man: Adaptive response to exercise. *J Physiol*, 270, 677-690.
- *Atomi, Y., & Miyashita, M. (1980). Effect of training intensity in adult females on aerobic power, related to lean body mass. *Eur J Appl Physiol*, 44, 109-116.
- *Bennett, B. L., Schlichting, C. L., & Bondi, K. R. (1985). Cardiorespiratory fitness and cognitive performance before and after confinement in a nuclear submarine. *Aviat Space Environ Med*, 56(11), 1085-1091.
- *Blaney, J., Sothmann, M., Raff, H., Hart, B., & Horn, T. (1990). Impact of exercise training on plasma adrenocorticotropin response to a well-learned vigilance task. *Psychoneuroendocrinology*, 15(5-6), 453-462.
- *Blessing, D. L., Wilson, G. D., Puckett, J. R., & Ford, H. T. (1987). The physiologic effects of eight weeks of aerobic dance with and without hand-held weights. *Am J Sports Med*, 15(5), 508-510.
- *Blumenthal, J. A., Emery, C. F., Walsh, M. A., Cox, D. R., Kuhn, C. M., Williams, R. B., & Williams, R. S. (1988). Exercise training in healthy type A middle-aged men: Effects on behavioral and cardiovascular responses. *Psychosom Med*, 50(4), 418-433.
- *Blumenthal, J. A., Fredrikson, M., Kuhn, C. M., Ulmer, R. L., Walsh-Riddle, M., & Appelbaum, M. (1990). Aerobic exercise reduces levels of cardiovascular and sympathoadrenal responses to mental stress in subjects without prior evidence of myocardial ischemia. *Am J Cardiol*, 65(1), 93-98.
- *Bonaduce, D., Petretta, M., Cavallaro, V., Apicella, C., Ianniciello, A., Romano, M., Breglio, R., & Marciano, F. (1998). Intensive training and cardiac autonomic control in high level athletes. *Med Sci Sports Exerc*, 30(5), 691-696.
- *Boreham, C. A., Kennedy, R. A., Murphy, M. H., Tully, M., Wallace, W. F., & Young, I. (2005). Training effects of short bouts of stair climbing on cardiorespiratory fitness, blood lipids, and homocysteine in sedentary young women. *Br J Sports Med*, 39(9), 590-593.
- *Boutcher, S. H., Seip, R. L., Hetzler, R. K., Pierce, E. F., Snead, D., & Weltman, A. (1989). The effects of specificity of training on rating of perceived exertion at the lactate threshold. *Eur J Appl Physiol Occup Physiol*, 59(5), 365-369.
- *Branch, J. D., Pate, R. R., Bourque, S. P., Convertino, V. A., Durstine, J. L., & Ward, D. S. (1997). Effects of exercise mode on hematologic adaptations to endurance training in adult females. *Aviat Space Environ Med*, 68(9), 788-794.

- *Branch, J. D., Pate, R. R., Bourque, S. P., Convertino, V. A., Durstine, J. L., & Ward, D. S. (1999). Exercise training and intensity does not alter vascular volume responses in women. *Aviat Space Environ Med*, 70(11), 1070-1076.
- *Bryner, R. W., Toffle, R. C., Ullrich, I. H., & Yeater, R. A. (1997). The effects of exercise intensity on body composition, weight loss, and dietary composition in women. *J Am Coll Nutr*, 16(1), 68-73.
- *Brynteson, P., & Sinning, W. E. (1973). The effects of training frequencies on the retention of cardiovascular fitness. *Med Sci Sports Exerc*, 5, 29-33.
- *Burke, E. J., & Franks, B. D. (1975). Changes in $\dot{V}_{O_{2max}}$ resulting from bicycle training *Res Q*, 46, 31-37.
- *Cadefau, J., Green, H. J., Cusso, R., Ball-Burnett, M., & Jamieson, G. (1994). Coupling of muscle phosphorylation potential to glycolysis during work after short-term training. *J Appl Physiol*, 76(6), 2586-2593.
- *Campbell, K. L., Westerlind, K. C., Harber, V. J., Bell, G. J., Mackey, J. R., & Courneya, K. S. (2007). Effects of aerobic exercise training on estrogen metabolism in premenopausal women: A randomized controlled trial. *Cancer Epidemiol Biomarkers Prev*, 16(4), 731-739.
- *Carter, H., Jones, A. M., & Doust, J. H. (1999). Effect of 6 weeks of endurance training on the lactate minimum speed. *J Sports Sci*, 17(12), 957-967.
- *Chtara, M., Chamari, K., Chaouachi, M., Chaouachi, A., Koubaa, D., Feki, Y., & Millet, G. P., Amri, M. (2005). Effects of intra-session concurrent endurance and strength training sequence on aerobic performance and capacity. *Br J Sports Med*, 39(8), 555-560.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- *Convertino, V. A., Mack, G. W., & Nadel, E. R. (1991). Elevated central venous pressure: A consequence of exercise training-induced hypervolemia? *Am J Physiol*, 260(2), R273-277.
- *Cristell, M., Hutchinson, K. M., & Alessio, H. M. (1998). Effects of exercise training on hearing ability. *Scand Audiol*, 27(4), 219-224.
- *Cunningham, D. A., McCrimmon, D., & Vlach, L. F. (1979). Cardiovascular response to interval and continuous training in women. *J Appl Physiol*, 41, 187-197.
- *Daussin, F. N., Zoll, J., Dufour, S. P., Ponsot, E., Lonsdorfer-Wolf, E., Doutreleau, S., Mettauer, B., Piquard, F., Geny, B., & Richard, R. (2008). Effect of interval versus continuous training on cardiorespiratory and mitochondrial functions: Relationship to aerobic performance improvements in sedentary subjects. *Am J Physiol Regul Integr Comp Physiol*, 295(1), R264-272.
- *Davis, J. A., Frank, M. H., Whipp, B. J., & Wasserman, K. (1979). Anaerobic threshold alterations caused by endurance training in middle-aged men. *J Appl Physiol*, 46(6), 1039-1046.
- *de Geus, E. J., Kluft, C., de Bart, A. C., & van Doornen, L. J. (1992). Effects of exercise training on plasminogen activator inhibitor activity. *Med Sci Sports Exerc*, 24(11), 1210-1219.
- *De Meersman, R. E. (1992). Respiratory sinus arrhythmia alteration following training in endurance athletes. *Eur J Appl Physiol Occup Physiol*, 64(5), 434-436.
- *Demarle, A. P., Heugas, A. M., Slawinski, J. J., Tricot, V. M., Koralsztejn, J. P., & Billat, V. L. (2003). Whichever the initial training status, any increase in velocity at lactate threshold appears as a major factor in improved time to exhaustion at the same severe velocity after training. *Arch Physiol Biochem*, 111(2), 167-176.

- *Denis, C., Dormois, D., & Lacour, J. R. (1984). Endurance training, $\dot{V}_{O_{2max}}$, and OBLA: A longitudinal study of two different age groups. *Int J Sports Med*, 5, 167-173.
- *Desharnais, R., Jobin, J., Cote, C., Levesque, L., Godin, G. (1993). Aerobic exercise and the placebo effect: A controlled study. *Psychosom Med*, 55(2), 149-154.
- *Despres, J. P., Moorjani, S., Tremblay, A., Poehlman, E. T., Lupien, P. J., Nadeau, A., & Bouchard, C. (1988). Heredity and changes in plasma lipids and lipoproteins after short-term exercise training in men. *Arterioscler Thromb Vasc Biol*, 8, 402-409.
- *Duey, W. J., O'Brien, W. L., Crutchfield, A. B., Brown, L. A., Williford, H. N., & Sharff-Olson, M. (1998). Effects of exercise training on aerobic fitness in African-American females. *Ethn Dis*, 8(3), 306-311.
- *Duffey, D. J., Horwitz, L. D., & Brammell, H. L. (1984). Nifedipine and the conditioning response. *Am J Cardiol*, 53(7), 908-911.
- *Egana, M., & Donne, B. (2004). Physiological changes following a 12-week gym-based stair-climbing, elliptical trainer and treadmill running program in females. *J Sports Med Phys Fitness*, 44(2), 141-146.
- *Esfarjani, F., & Laursen, P. B. (2007). Manipulating high-intensity interval training: Effects on $\dot{V}_{O_{2max}}$, the lactate threshold, and 3000 m running performance in moderately trained males. *J Sci Med Sport*, 10(1), 27-35.
- *Fox, E. L., Bartels, R. L., Billings, C. E., Mathews, D. K., Bason, R., & Webb, W. M. (1973). Intensity and distance of interval training programs and changes in aerobic power. *Med Sci Sports Exerc*, 5, 18-22.
- *Fox, E. L., Bartels, R. L., Billings, C. E., O'Brien, R., Bason, R., & Mathews, D. K. (1975). Frequency and duration of interval training programs and changes in aerobic power. *J Appl Physiol*, 38, 481-484.
- *Fox, E. L., Bartels, R. L., Klinzing, J., & Ragg, K. (1977). Metabolic responses to interval training programs of high and low power output. *Med Sci Sports Exerc*, 9, 191-196.
- *Freund, B. J., Allen, D., & Wilmore, J. H. (1986). Interaction of test protocol and inclined run training on maximal oxygen uptake. *Med Sci Sports Exerc*, 18(5), 588-592.
- *Fringer, M. N., & Stull, G. A. (1974). Changes in cardiorespiratory parameters during periods of training and detraining in young adult females. *Med Sci Sports* 6, 20-25.
- *Gaesser, G. A., & Rich, R. G. (1984). Effects of high- and low-intensity exercise training on aerobic capacity and blood lipids. *Med Sci Sports Exerc*, 16, 269-274.
- *Gaiga, M. C., & Docherty, D. (1995). The effect of an aerobic interval training program on intermittent anaerobic performance. *Can J Appl Physiol*, 20(4), 452-464.
- *Gaine, P. C., Viesselman, C. T., Pikosky, M. A., Martin, W. F., Armstrong, L. E., Pescatello, L. S., & Rodriguez, N. R. (2005). Aerobic exercise training decreases leucine oxidation at rest in healthy adults. *J Nutr*, 135(5), 1088-1092.
- *Gallo Junior, L., Maciel, B. C., Marin-Neto, J. A., & Martins, L. E. (1989). Sympathetic and parasympathetic changes in heart rate control during dynamic exercise induced by endurance training in man. *Braz J Med Biol Res*, 22(5), 631-643.
- *Gamble, R. P., Boreham, C. A., & Stevens, A. B. (1993). Effects of a 10-week exercise intervention programme on exercise and work capacities in Belfast's ambulance-men. *Occup Med (Lond)*, 43(2), 85-89.
- *Garber, C. E., McKinney, J. S., & Carleton, R. A. (1992). Is aerobic dance an effective alternative to walk-jog exercise training? *J Sports Med Phys Fitness*, 32(2), 136-141.
- *Gettman, L. R., Pollock, M. L., Durstine, J. L., Ward, A., Ayres, J., & Linnerud, A. C. (1977).

- Physiological responses of men to 1, 3, and 5 day per week training programs. *Res Q*, 47, 638-646.
- *Gettman, L. R., Ayres, J. J., Pollock, M. L., & Jackson, A. (1978). The effect of circuit weight training on strength, cardiorespiratory function, and body composition of adult men. *Med Sci Sports*, 10(3), 171-176.
- *Gibbons, E. S., Jessup, G. T., Wells, T. D., & Werthmann, D. A. (1983). Effects of various training intensity levels on anaerobic threshold and aerobic capacity in females. *J Sports Med Phys Fit*, 23, 315-318.
- *Gilders, R. M., Voner, C., & Dudley, G. A. (1989). Endurance training and blood pressure in normotensive and hypertensive adults. *Med Sci Sports Exerc*, 21, 629-636.
- *Golden, H. P., & Vaccaro, P. (1984). The effects of endurance training on the anaerobic threshold. *J Sports Med Phys Fit*, 24, 205-211.
- *Gormley, S. E., Swain, D. P., High, R., Spina, R. J., Dowling, E. A., Kotipalli, U. S., & Gandrakota, R. (2008). Effect of Intensity of Aerobic Training on $\dot{V}_{O_{2max}}$. *Med Sci Sports Exerc*, 40(7), 1336-1343.
- *Gregory, L. W. (1979). The development of aerobic capacity: A comparison of continuous and interval training. *Res Q*, 50, 199-206.
- *Haennel, R., Teo, K. K., Quinney, A., & Kappagoda, T. (1989). Effects of hydraulic circuit training on cardiovascular function. *Med Sci Sports Exerc*, 21(5), 605-612.
- *Hagan, R. D., Upton, S. J., Wong, L., & Whittam, J. (1986). The effect of aerobic conditioning and/or calorie restriction in overweight men and women. *Med Sci Sports Exerc*, 18, 87-94.
- *Hakkinen, K. (1993). Changes in physical fitness profile in female volleyball players during the competitive season. *J Sports Med Phys Fitness*, 33(3), 223-232.
- *Hass, C. J., Garzarella, L., de Hoyos, D. V., Connaughton, D. P., & Pollock, M. L. (2001). Concurrent improvements in cardiorespiratory and muscle fitness in response to total body recumbent stepping in humans. *Eur J Appl Physiol*, 85(1-2), 157-163.
- Hays, W. L. (1963). *Statistics for Psychologists*. New York: Holt, Rinehart, Winston.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- *Helgerud, J., Engen, L. C., Wisloff, U., & Hoff, J. (2001). Aerobic endurance training improves soccer performance. *Med Sci Sports Exerc*, 33(11), 1925-1931.
- *Helgerud, J., Hoydal, K., Wang, E., Karlsen, T., Berg, P., Bjerkaas, M., ... Hoff, J. (2007). Aerobic high-intensity intervals improve $\dot{V}_{O_{2max}}$ more than moderate training. *Med Sci Sports Exerc*, 39(4), 665-671.
- *Hendriksen, I. J., Zuiderveld, B., Kemper, H. C., & Bezemer, P. D. (2000). Effect of commuter cycling on physical performance of male and female employees. *Med Sci Sports Exerc*, 32(2), 504-510.
- *Henritze, J., Weltman, A., Schurrer, R. L., & Barlow, K. (1985). Effects of training at and above the lactate threshold and maximal oxygen uptake. *Eur J Appl Physiol*, 54, 84-88.
- *Hewitt, J. A., Whyte, G. P., Moreton, M., van Someren, K. A., & Levine, T. S. (2008). The effects of a graduated aerobic exercise programme on cardiovascular disease risk factors in the NHS workplace: A randomised controlled trial. *J Occup Med Toxicol*, 3, 7.
- Heyward, V. H. (1997). *Advanced Fitness Assessment and Exercise Prescription* (3rd ed.). Champaign, IL: Human Kinetics Publishers.
- *Hiatt, W. R., Marsh, R. C., Brammell, H. L., Fee, C., & Horwitz, L. D. (1984). Effect of aerobic conditioning on the peripheral circulation during chronic beta-adrenergic blockade. *J Am*

- Coll Cardiol*, 4(5), 958-963.
- *Hickson, R. C., Bomze, H. A., & Holloszy, J. O. (1977). Linear increase in aerobic power induced by a strenuous program of endurance exercise. *J Appl Physiol*, 42, 372-376.
 - *Hickson, R. C., Foster, C., Pollock, M. L., Galassi, T. M., & Rich, S. (1985). Reduced training intensities and loss of aerobic power, endurance, and cardiac growth. *J Appl Physiol*, 58, 492-499.
 - *Hoppeler, H., Howald, H., Conley, K., Lindstedt, S. L., Claassen, H., Vock, P., & Weibel, E. R. (1985). Endurance training in humans: Aerobic capacity and structure of skeletal muscle. *J Appl Physiol*, 59, 320-327.
 - *Houmard, J. A., Costill, D. L., Mitchell, J. B., Park, S. H., & Chenier, T. C. (1991). The role of anaerobic ability in middle distance running performance. *Eur J Appl Physiol Occup Physiol*, 62(1), 40-43.
 - *Hurley, B. F., Hagberg, J. M., Allen, W. K., Seals, D. R., Young, J. C., Cuddihee, R. W., & Holloszy, J. O. (1984). Effect of training on blood lactate levels during submaximal exercise. *J Appl Physiol*, 56, 1260-1264.
 - *Hurley, B. F., Seals, D. R., Ehsani, A. A., Cartier, L. J., Dalsky, G. P., Hagberg, J. M., & Holloszy, J. O. (1984). Effects of high-intensity strength training on cardiovascular function. *Med Sci Sports Exerc*, 16, 483-485.
 - *Iaia, F. M., Hellsten, Y., Nielsen, J. J., Fernstrom, M., Sahlin, K., & Bangsbo, J. (2009). Four weeks of speed endurance training reduces energy expenditure during exercise and maintains muscle oxidative capacity despite a reduction in training volume. *J Appl Physiol*, 106(1), 73-80.
 - *Impellizzeri, F. M., Rampinini, E., Maffiuletti, N. A., Castagna, C., Bizzini, M., & Wisloff, U. (2008). Effects of aerobic training on the exercise-induced decline in short-passing ability in junior soccer players. *Appl Physiol Nutr Metab*, 33(6), 1192-1198.
 - *Ingjer, F., & Myhre, K. (1992). Physiological effects of altitude training on elite male cross-country skiers. *J Sports Sci*, 10(1), 37-47.
 - *Jenkins, D. G., & Quigley, B. M. (1992). Endurance training enhances critical power. *Med Sci Sports Exerc*, 24(11), 1283-1289.
 - *Jenkins, D. G., & Quigley, B. M. (1993). The influence of high-intensity exercise training on the Wlim-Tlim relationship. *Med Sci Sports Exerc*, 25(2), 275-282.
 - *Jette, M., Mongeon, J., & Shephard, R. J. (1982). Demonstration of a training response by the Canadian home fitness test. *Eur J Appl Physiol Occup Physiol*, 49(2), 143-150.
 - *Kaikkonen, H., Yrjama, M., Siljander, E., Byman, P., & Laukkanen, R. (2000). The effect of heart rate controlled low resistance circuit weight training and endurance training on maximal aerobic power in sedentary adults. *Scand J Med Sci Sports*, 10(4), 211-215.
 - *Katayama, K., Sato, Y., Ishida, K., Mori, S., & Miyamura, M. (1998). The effects of intermittent exposure to hypoxia during endurance exercise training on the ventilatory responses to hypoxia and hypercapnia in humans. *Eur J Appl Physiol Occup Physiol*, 78(3), 189-194.
 - *Katayama, K., Shimoda, M., Maeda, J., & Takemiya, T. (1998). Endurance exercise training increases peripheral vascular response in human fingers. *Jpn J Physiol*, 48(5), 365-371.
 - *Kiens, B., Jorgensen, I., Lewis, S. M., Jensen, G., Lithell, H., Vessby, B., Hoe, S., & Schnohr, P. (1980). Increased plasma HDL-cholesterol and apo A-1 in sedentary middle-aged men after physical conditioning. *Eur J Clin Invest*, 10(3), 203-209.
 - *Kim, H. S., Tanaka, K., & Maeda, K. (1991). Effects of exercise training at an intensity relative to lactate threshold in mildly obese women. *Ann Physiol Anthropol*, 10, 229-236.

- *Kiviniemi, A. M., Hautala, A. J., Kinnunen, H., & Tulppo, M. P. (2007). Endurance training guided individually by daily heart rate variability measurements. *Eur J Appl Physiol*, *101*(6), 743-751.
- *Knuttgen, H. G., Nordesjo, L. O., Ollander, B., & Saltin, B. (1973). Physical conditioning through interval training with young adult males. *Med Sci Sports*, *5*, 220-226.
- *Kraemer, W. J., Fleck, S. J., Callister, R., Shealy, M., Dudley, G. A., Maresh, ... Falkel, J. E. (1989). Training responses of plasma beta-endorphin, adrenocorticotropin, and cortisol. *Med Sci Sports Exerc*, *21*(2), 146-153.
- *Kraemer, W. J., Patton, J. F., Gordon, S. E., Harman, E. A., Deschenes, M. R., Reynolds, K., ... Dziados, J. E. (1995). Compatibility of high-intensity strength and endurance training on hormonal and skeletal muscle adaptations. *J Appl Physiol*, *78*(3), 976-989.
- *Krotkiewski, M., Mandroukas, K., Sjostrom, L., Sullivan, L., Wetterqvist, H., & Bjorntorp, P. (1979). Effects of long-term physical training on body fat, metabolism, and blood pressure in obesity. *Metabolism*, *28*, 650-658.
- *Lake, M. J., & Cavanagh, P. R. (1996). Six weeks of training does not change running mechanics or improve running economy. *Med Sci Sports Exerc*, *28*(7), 860-869.
- *Landry, F., Bouchard, C., & Dumesnil, J. (1985). Cardiac dimension changes with endurance training. Indications of a genotype dependency. *JAMA*, *254*(1), 77-80.
- *Lavoie, N. F. (1977). The aerobic cost of exercise as a determinant of endurance fitness. *Can J Appl Sports Sci*, *2*, 121-125.
- *Leddy, J. J., Limprasertkul, A., Patel, S., Modlich, F., Buyea, C., Pendergast, D. R., & Lundgren, C. E. (2007). Isocapnic hyperpnea training improves performance in competitive male runners. *Eur J Appl Physiol*, *99*(6), 665-676.
- *LeMura, L. M., von Duvillard, S. P., Andreacci, J., Klebez, J. M., Chelland, S. A., & Russo, J. (2000). Lipid and lipoprotein profiles, cardiovascular fitness, body composition, and diet during and after resistance, aerobic and combination training in young women. *Eur J Appl Physiol*, *82*(5-6), 451-458.
- *Lesmes, G. F., Fox, E. L., Stevens, C., & Otto, R. (1978). Metabolic responses of females to high intensity interval training of different frequencies. *Med Sci Sports*, *10*, 229-232.
- *Lieber, D. C., Lieber, R. L., & Adams, W. C. (1989). Effects of run-training and swim-training at similar absolute intensities on treadmill $\dot{V}_{O_{2max}}$. *Med Sci Sports Exerc*, *21*(6), 655-661.
- Light, R. J., & Pillemer, D. B. (1984). *Summing up: the science of reviewing research*. Cambridge, MA: Harvard University Press.
- *Lightfoot, J. T., Claytor, R. P., Torok, D. J., Journell, T. W., & Fortney, S. M. (1989). Ten weeks of aerobic training do not affect lower body negative pressure response. *J Appl Physiol*, *67*, 894-901.
- *Loimaala, A., Huikuri, H., Oja, P., Pasanen, M., & Vuori, I. (2000). Controlled 5-mo aerobic training improves heart rate but not heart rate variability or baroreflex sensitivity. *J Appl Physiol*, *89*(5), 1825-1829.
- *Loy, S. F., Holland, G. J., Mutton, D. L., Snow, J., Vincent, W. J., Hoffmann, J. J., & Shaw, S. (1993). Effects of stair-climbing vs. run training on treadmill and track running performance. *Med Sci Sports Exerc*, *25*(11), 1275-1278.
- Londoree, B. R. (1997). Effect of training on lactate/ventilatory thresholds: A meta-analysis. *Med Sci Sports Exerc*, *29*(6), 837-843.
- *Mack, G. W., Thompson, C. A., Doerr, D. F., Nadel, E. R., & Convertino, V. A. (1991). Diminished baroreflex control of forearm vascular resistance following training. *Med Sci*

- Sports Exerc*, 23, 1367-1374.
- *Mack, G. W., Convertino, V. A., & Nadel, E. R. (1993). Effect of exercise training on cardiopulmonary baroreflex control of forearm vascular resistance in humans. *Med Sci Sports Exerc*, 25(6), 722-726.
 - *Magazanik, A., Weinstein, Y., Abarbanel, J., Lewinski, U., Shapiro, Y., Inbar, O., & Epstein, S. (1991). Effect of an iron supplement on body iron status and aerobic capacity of young training women. *Eur J Appl Physiol Occup Physiol*, 62(5), 317-323.
 - *Magel, J. R., McArdle, W. D., Toner, M., & Delio, D. J. (1978). Metabolic and cardiovascular adjustment to arm training. *J Appl Physiol*, 45(1), 75-79.
 - *Malek, M. H., Housh, T. J., Coburn, J. W., Beck, T. W., Schmidt, R. J., Housh, D. J., & Johnson, G. O. (2006). Effects of eight weeks of caffeine supplementation and endurance training on aerobic fitness and body composition. *J Strength Cond Res*, 20(4), 751-755.
 - *Margaritis, I., Tessier, F., Prou, E., Marconnet, P., & Marini, J. F. (1997). Effects of endurance training on skeletal muscle oxidative capacities with and without selenium supplementation. *J Trace Elem Med Biol*, 11(1), 37-43.
 - *Marsh, R. C., Hiatt, W. R., Brammell, H. L., & Horwitz, L. D. (1983). Attenuation of exercise conditioning by low dose beta-adrenergic receptor blockade. *J Am Coll Cardiol*, 2(3), 551-556.
 - *McArdle, W. D., Magel, J. R., Delio, D. J., Toner, M., & Chase, J. M. (1978). Specificity of run training on $\dot{V}_{O_{2max}}$ and heart rate changes during running and swimming. *Med Sci Sports*, 10, 16-20.
 - *McCarthy, J. P., Agre, J. C., Graf, B. K., Pozniak, M. A., & Vailas, A. C. (1995). Compatibility of adaptive responses with combining strength and endurance training. *Med Sci Sports Exerc*, 27(3), 429-436.
 - *McDonald, M. P., Sanfilippo, A. J., & Savard, G. K. (1993). Baroreflex function and cardiac structure with moderate endurance training in normotensive men. *J Appl Physiol*, 74(5), 2469-2477.
 - *McDowell, S. L., Hughes, R. A., Hughes, R. J., Housh, T. J., & Johnson, G. O. (1992). The effect of exercise training on salivary immunoglobulin A and cortisol responses to maximal exercise. *Int J Sports Med*, 13(8), 577-580.
 - *McLellan, T. M., & Skinner, J. S. (1981). The use of the aerobic threshold as a basis for training. *Can J Appl Sports Sci*, 6, 197-201.
 - *Melanson, E. L., Freedson, P. S., & Jungbluth, S. (1996). Changes in $\dot{V}_{O_{2max}}$ and maximal treadmill time after 9 wk of running or in-line skate training. *Med Sci Sports Exerc*, 28(11), 1422-1426.
 - *Meredith, C. N., Frontera, W. R., Fisher, E. C., Hughes, V. A., Herland, J. C., Edwards, J., & Evans, W. J. (1989). Peripheral effects of endurance training in young and old subjects. *J Appl Physiol*, 66, 2844-2849.
 - *Meyer, T., Auracher, M., Heeg, K., Urhausen, A., & Kindermann, W. (2007). Effectiveness of low-intensity endurance training. *Int J Sports Med*, 28(1), 33-39.
 - *Mikesell, K. A., & Dudley, G. A. (1984). Influence of intense endurance training on aerobic power of competitive distance runners. *Med Sci Sports Exerc*, 16, 371-375.
 - *Milburn, S., & Butts, N. K. (1983). A comparison of the training responses to aerobic dance and jogging in college females. *Med Sci Sports Exerc*, 15(6), 510-513.
 - *Milesis, C. A., Pollock, M. L., Bah, M. D., Ayres, A. W., & Linnerud, A. C. (1976). Effects of different durations of physical training on cardiorespiratory function, body composition,

- and serum lipids. *Res Q*, 47, 716-725.
- *Millet, G. P., Jaouen, B., Borrani, F., & Candau, R. (2002). Effects of concurrent endurance and strength training on running economy and $\dot{V}O_2$ kinetics. *Med Sci Sports Exerc*, 34(8), 1351-1359.
 - *Miyachi, M., Tanaka, H., Yamamoto, K., Yoshioka, A., Takahashi, K., & Onodera, S. (2001). Effects of one-legged endurance training on femoral arterial and venous size in healthy humans. *J Appl Physiol*, 90(6), 2439-2444.
 - *Moffatt, R. J., Stamford, B. A., & Neill, R. D. (1977). Placement of tri-weekly training sessions: Importance regarding enhancement of aerobic capacity. *Res Q*, 48, 583-589.
 - Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1), 105-125.
 - *Mourot, L., Perrey, S., Tordi, N., & Rouillon, J. D. (2004). Evaluation of fitness level by the oxygen uptake efficiency slope after a short-term intermittent endurance training. *Int J Sports Med*, 25(2), 85-91.
 - *Murphy, M. H., & Hardman, A. E. (1998). Training effects of short and long bouts of brisk walking in sedentary women. *Med Sci Sports Exerc*, 30(1), 152-157.
 - *Murray, A., Delaney, T., & Bell, C. (2006). Rapid onset and offset of circulatory adaptations to exercise training in men. *J Hum Hypertens*, 20(3), 193-200.
 - *Mutton, D. L., Loy, S. F., Rogers, D. M., Holland, G. J., Vincent, W. J., & Heng, M. (1993). Effect of run vs combined cycle/run training on $\dot{V}O_{2max}$ and running performance. *Med Sci Sports Exerc*, 25(12), 1393-1397.
 - *O'Connor, P. J., Petruzzello, S. J., Kubitz, K. A., & Robinson, T. L. (1995). Anxiety responses to maximal exercise testing. *Br J Sports Med*, 29(2), 97-102.
 - *Oja, P., Laukkanen, R. M., Kukkonen-Harjula, T. K., Vuori, I. M., Pasanen, M. E., Niittymaki, S. P., & Solakivi, T. (1991). Training effects of cross-country skiing and running on maximal aerobic cycle performance and on blood lipids. *Eur J Appl Physiol Occup Physiol*, 62(6), 400-404.
 - *Panton, L. B., Guillen, G. J., Williams, L., Graves, J. E., Vivas, C., Cediell, M., Pollock, M. L., Garzarella, L., Krumerman, J., & Derendorf, H. (1995). The lack of effect of aerobic exercise training on propranolol pharmacokinetics in young and elderly adults. *J Clin Pharmacol*, 35(9), 885-894.
 - *Pawelczyk, J. A., Kenney, W. L., & Kenney, P. (1988). Cardiovascular responses to head-up tilt after an endurance exercise program. *Aviat Space Environ Med*, 59(2), 107-112.
 - *Pierce, E. F., Weltman, A., Seip, R. L., & Snead, D. (1990). Effects of training specificity on the lactate threshold and $\dot{V}O_2$ peak. *Int J Sports Med*, 11, 267-272.
 - *Pollock, M. L., Cureton, T. K., & Greninger, L. (1969). Effects of frequency of training on working capacity, cardiovascular function, and body composition of adult men. *Med Sci Sports*, 4, 6-10.
 - *Pollock, M. L., Dimmick, J., Miller, H. S., Kendrick, Z., & Linnerud, A. C. (1975). Effects of mode of training on cardiovascular function and body composition of adult men. *Med Sci Sports*, 7, 139-145.
 - *Prud'Homme, D., Bouchard, C., Leblanc, C., Landry, F., & Fontaine, E. (1984). Sensitivity of maximal aerobic power to training is genotype-dependent. *Med Sci Sports Exerc*, 16, 489-493.
 - *Quinn, T. J., Klooster, J. R., & Kenefick, R. W. (2006). Two short, daily activity bouts vs. one long bout: are health and fitness improvements similar over twelve and twenty-four weeks?

- J Strength Cond Res*, 20(1), 130-135.
- R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- *Rahnama, N., Gaeini, A. A., & Hamedinia, M. R. (2007). Oxidative stress responses in physical education students during 8 weeks aerobic training. *J Sports Med Phys Fitness*, 47(1), 119-123.
- *Ramsbottom, R., Williams, C., Fleming, N., & Nute, M. L. G. (1989). Training induced physiological and metabolic changes associated with improvements in running performance. *Br J Sports Med*, 23, 171-176.
- *Ramsbottom, R., Nevill, A. M., Seager, R. D., & Hazeldine, R. (2001). Effect of training on accumulated oxygen deficit and shuttle run performance. *J Sports Med Phys Fitness*, 41(3), 281-290.
- *Ray, C. A., & Cureton, K. J. (1991). Interactive effects of body posture and exercise training on maximal oxygen uptake. *J Appl Physiol*, 71(2), 596-600.
- *Ready, A. E., Eynon, R. B., & Cunningham, D. A. (1981). Effect of interval training and detraining on anaerobic fitness in women. *Can J Appl Sport Sci*, 6(3), 114-118.
- *Ready, A. E., & Quinney, H. A. (1982). Alterations in anaerobic threshold as the result of endurance training and detraining. *Med Sci Sports Exerc*, 14, 292-296.
- *Ready, A. E., & Quinney, H. A. (1982). The response of serum lipids and lipoproteins to high intensity endurance training. *Can J Appl Sport Sci*, 7(3), 202-208.
- *Roelen, C. A., de Vries, W. R., Koppeschaar, H. P., Vervoorn, C., Thijssen, J. H., & Blankenstein, M. A. (1997). Plasma insulin-like growth factor-I and high affinity growth hormone-binding protein levels increase after two weeks of strenuous physical training. *Int J Sports Med*, 18(4), 238-241.
- Rosenthal, R. (1979). The 'file-drawer problem' and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research*. New York: McGraw-Hill.
- *Rumley, A. G., Taylor, R., Grant, S., Pettigrew, A. R., Findlay, I., & Dargie, H. (1983). Effect of marathon training on the plasma lactate response to submaximal exercise in middle-aged men. *Br J Sports Med*, 22, 31-34.
- *Rusko, H., & Bosco, C. C. (1987). Metabolic response of endurance athletes to training with added load. *Eur J Appl Physiol Occup Physiol*, 56(4), 412-418.
- *Sagiv, M., & Ben-Sira, D. (2007). Weight lifting training and left ventricular function in adolescent subjects. *J Sports Med Phys Fitness*, 47(3), 329-334.
- *Saltin, B., Hartley, L. H., Kilbom, A., & Astrand, I. (1969). Physical training in sedentary middle-aged and older men. *Scan J Clin Lab Invest*, 24, 323-324.
- Samitz, G., & Bachl, N. (1991). Physical training programs and their effects on aerobic capacity and coronary risk profile in sedentary individuals: Design of a long-term exercise training program. *J Sports Med Phys Fitness*, 31(2), 283-293.
- *Santiago, M. C., Leon, A. S., & Serfass, R. C. (1995). Failure of 40 weeks of brisk walking to alter blood lipids in normolipemic women. *Can J Appl Physiol*, 20(4), 417-428.
- *Santtila, M., Keijo, H., Laura, K., & Heikki, K. (2008). Changes in cardiovascular performance during an 8-week military basic training period combined with added endurance or strength training. *Mil Med*, 173(12), 1173-1179.

- *Sato, Y., Hayamizu, S., Yamamoto, C., Ohkuwa, Y., Yamanouchi, K., & Sakamoto, N. (1986). Improved insulin sensitivity in carbohydrate and lipid metabolism after physical training. *Int J Sports Med*, 7(6), 307-310.
- *Sjodin, B., Jacobs, I., & Svedenhag, J. (1982). Changes in onset of blood lactate accumulation (OBLA) and muscle enzymes after training at OBLA. *Eur J Appl Physiol*, 49, 45-57.
- *Smith, D. A., & O'Donnell, T. V. (1984). The time course during 36 week's endurance training of $\dot{V}_{O_{2max}}$ and anaerobic threshold as determined with a new computerized method. *Clin Sci*, 67, 229-236.
- *Smith, D. J., & Wenger, H. A. (1981). The 10 day aerobic mini-cycle: The effects of interval or continuous training at two different intensities. *J Sports Med Phys Fit*, 21, 390-394.
- *Smith, T. P., Coombes, J. S., & Geraghty, D. P., (2003). Optimising high-intensity treadmill training using the running speed at maximal O₂ uptake and the time for which this can be maintained. *Eur J Appl Physiol*, 89(3-4), 337-343.
- *Spate-Douglas, T., & Keyser, R. E. (1999). Exercise intensity: Its effect on the high-density lipoprotein profile. *Arch Phys Med Rehabil*, 80(6), 691-695.
- *Sporis, G., Ruzic, L., & Leko, G. (2008). Effects of a new experimental training program on $\dot{V}_{O_{2max}}$ and running performance. *J Sports Med Phys Fitness*, 48(2), 158-165.
- *Svedenhag, J., Henriksson, J., Juhlin-Dannfelt, A., & Asano, K., (1984). Beta-adrenergic blockade and training in healthy men—effects on central circulation. *Acta Physiol Scand*, 120(1), 77-86.
- *Sweeney, M. E., Fletcher, B. J., & Fletcher, G. F. (1989). Exercise testing and training with beta-adrenergic blockade: Role of the drug washout period in "unmasking" a training effect. *Am Heart J*, 118(5 Pt 1), 941-946.
- *Takeno, Y., Kamijo, Y. I., & Nose, H. (2001). Thermoregulatory and aerobic changes after endurance training in a hypobaric hypoxic and warm environment. *J Appl Physiol*, 91(4), 1520-1528.
- *Tanaka, K., Watanabe, H., Konishi, Y., Mitsuzono, R., Sumida, S., Tanaka, S., Fukuda, T., & Nakadomo, F. (1986). Longitudinal associations between anaerobic threshold and distance running performance. *Eur J Appl Physiol*, 55, 248-252.
- *Tessier, F., Margaritis, I., Richard, M. J., Moynot, C., & Marconnet, P. (1995). Selenium and training effects on the glutathione system and aerobic performance. *Med Sci Sports Exerc*, 27(3), 390-396.
- *Thomas, T. R., Adeniran, S. B., & Etheridge, G. L. (1984). Effects of different running programs on $\dot{V}_{O_{2max}}$, percent fat, and plasma lipids. *Can J Appl Sports Sci*, 9, 55-62.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
- *van der Woude, L. H., van Croonenborg, J. J., Wolff, I., Dallmeijer, A. J., & Hollander, A. P. (1999). Physical work capacity after seven weeks of wheelchair training: Effect of intensity in able-bodied subjects. *Med Sci Sports Exerc*, 31(2), 331-341.
- *Wang, J. S., Jen, C. J., & Chen, H. I. (1995). Effects of exercise training and deconditioning on platelet function in men. *Arterioscler Thromb Vasc Biol*, 15(10), 1668-1674.
- *Warburton, D. E., Haykowsky, M. J., Quinney, H. A., Blackmore, D., Teo, K. K., Taylor, D. A., McGavock, J., & Humen, D. P. (2004). Blood volume expansion and cardiorespiratory function: Effects of training modality. *Med Sci Sports Exerc*, 36(6), 991-1000.
- *Watson, R. R., Moriguchi, S., Jackson, J. C., Werner, L., Wilmore, J. H., & Freund, B. J. (1986).

- Modification of cellular immune functions in humans by endurance exercise training during beta-adrenergic blockade with atenolol or propranolol. *Med Sci Sports Exerc*, 18(1), 95-100.
- *Wehrlin, J. P., Zuest, P., Hallen, J., & Marti, B. (2006). Live high-train low for 24 days increases hemoglobin mass and red cell volume in elite endurance athletes. *J Appl Physiol*, 100(6), 1938-1945.
- *Weir, L. L., Weir, J. P., Housh, T. J., & Johnson, G. O. (1997). Effect of an aerobic training program on physical working capacity at heart rate threshold. *Eur J Appl Physiol Occup Physiol*, 75(4), 351-356.
- *Weltman, A., Seip, R. L., Snead, D., Weltman, J. Y., Haskvitz, E. M., Evans, W. S., Veldhuis, J. D., & Rogol, A. D. (1992). Exercise training at and above the lactate threshold in previously untrained women. *Int J Sports Med*, 13, 257-263.
- Wenger, H. A., & Bell, G.J. (1986). The interactions of intensity, frequency and duration of exercise training in altering cardiorespiratory fitness. *Sports Med*, 3(5), 346-356.
- *White, L. J., Dressendorfer, R. H., Ferguson, M. A., & Wade, C. E. (2002). Maintenance of testosterone status in fitness joggers after increased training mileage. *Eur J Appl Physiol*, 86(6), 498-502.
- *Wieling, W., Borghols, E. A., Hollander, A. P., Danner, S. A., & Dunning, A. J. (1981). Echocardiographic dimensions and maximal oxygen uptake in oarsmen during training. *Br Heart J*, 46(2), 190-195.
- *Wilcox, A. R., & Bulbulian, R. (1984). Changes in running economy to $\dot{V}_{O_{2max}}$ during a cross-country season. *J Sports Med*, 24, 321-327.
- *Williford, H. N., Richards, L. A., Scharff-Olson, M., Brown, J., Blessing, D., & Duey, W. J. (1998). Bench stepping and running in women. Changes in fitness and injury status. *J Sports Med Phys Fitness*, 38(3), 221-226.
- *Wilmore, J. H., Royce, J., Girandola, R. N., Katch, F. I., & Katch, V. L. (1970). Physiological alterations resulting from a 10-week program of jogging. *Med Sci Sports Exerc*, 2, 7-14.
- *Wilmore, J. H., Davis, J. A., O'Brien, R. S., Vodak, P. A., Walder, G. R., & Amsterdam, E. A. (1980). Physiological alterations consequent to 20-week conditioning programs of bicycling, tennis, and jogging. *Med Sci Sports Exerc*, 12, 1-8.
- *Wirth, J. C., Lohman, T. G., Avallone, J. P., Shire, T., & Boileau, R. A. (1978). The effect of physical training on the serum iron levels of college-age women. *Med Sci Sports*, 10(3), 223-226.
- *Wolfe, L. A., Laprade, A., Burggraf, G. W., & Norman, R. (1992). Cardiac responses of young women to conditioning for a 10 km race. *Int J Sports Med*, 13(5), 384-389.
- *Wolfel, E. E., Hiatt, W. R., Brammell, H. L., Carry, M. R., Ringel, S. P., Travis, V., & Horwitz, L. D. (1986). Effects of selective and nonselective beta-adrenergic blockade on mechanisms of exercise conditioning. *Circulation*, 74(4), 664-674.
- *Woods, N. G., & Graham, T. E. (1986). Effect of menstrual cycle phase and exercise training on serum lipids. *Can J Appl Sport Sci*, 11(2), 88-93.
- *Yoshida, T., Suda, Y., & Takeuchi, N. (1982). Endurance training regimen based upon arterial blood lactate: Effects on anaerobic threshold. *Eur J Appl Physiol*, 49, 223-230.
- *Yoshida, T., Udo, M., Chida, M., Ichioka, M., Makiguchi, K., & Yamaguchi, T. (1990). Specificity of physiological adaptation to endurance training in distance runners and competitive walkers. *J Appl Physiol*, 61, 197-201.

*Zapico, A. G., Calderon, F. J., Benito, P. J., Gonzalez, C. B., Parisi, A., Pigozzi, F., & Di Salvo, V. (2007). Evolution of physiological and haematological parameters with training load in elite male road cyclists: A longitudinal study. *J Sports Med Phys Fitness*, 47(2), 191-196.

