Open access • Journal Article • DOI:10.1007/S40279-017-0680-8

Inter-Individual Responses of Maximal Oxygen Uptake to Exercise Training: A Critical Review — Source link

Philip Williamson, Greg Atkinson, Alan M. Batterham

Institutions: Teesside University

Published on: 17 Jan 2017 - Sports Medicine (Springer International Publishing)

Williamson, P.J. Atkinson, G. Batterham, A.M.

Inter-Individual Responses of Maximal Oxygen Uptake to Exercise Training: A Critical Review

Health and Social Care Institute

Teesside University

Middlesbrough

UK

TS1 3BX

Corresponding Author: Williamson, P.J. (Health and Social Care Institute, Teesside University, Middlesbrough, UK, TS1 3BX, email: P.Williamson@tees.ac.uk, phone: 01642384143)

Running title: Inter-Individual VO_2max Response to Exercise: A Critical Review

Abstract

It has recently been reported how to quantify inter-individual differences in the response to an exercise intervention using the standard deviation (SD) of the change scores, as well as how to appraise these differences for clinical relevance. In a parallel-group RCT, the key trigger for further investigation into inter-individual responses is when the standard deviation of change (SD_{change}) in the intervention sample is substantially larger than the same standard deviation derived from a suitable comparator sample. 'True' and clinically relevant inter-individual differences in response can then be plausibly expected, and potential moderators and mediators of the inter-individual differences can be explored. We now aim to critically review the research on the inter – individual differences in response to exercise training, focusing on maximal oxygen uptake (VO_2max).

A literature search through the relevant bibliographic databases resulted in the identification of six relevant studies that were published prior to the influential HERITAGE family study. Only one of these studies was found to include a comparator arm. Re-analysis of the data from this study, accounting for random within-subjects variation, revealed an absence of clinically important inter-individual differences in the response of VO_2max to exercise training. The SD_{change} was, in fact, larger (± 5.6 ml/kg/min) for the comparator than the intervention group (± 3.7 ml/kg/min).

We located over 180 publications that resulted from the HERITAGE study, but we could not find a comparator arm in any of these studies. Some authors did not explain this absence, while others reasoned that only inter-individual differences in exercise response were of interest, so this intervention sample was investigated solely. We also found this absence of a comparator sample in on-going studies. A perceived high test-retest reliability is offered as a justification for the absence of a comparator arm, but the test-retest reliability analysis for HERITAGE was over a much shorter term than the length of the actual training period between baseline and follow-up measurements of VO_2max.

We also scrutinised the studies in which twins have been investigated, resulting in concerns about how genetic influences on the magnitude of general within-subjects variability has been partitioned out (again in the absence of a comparator – no training – group), as well as with the intraclass correlation coefficient approach to data analysis. Twin pairs were found to be sometimes heterogeneous for the obviously influential factors of sex, age and fitness, thereby inflating an unadjusted coefficient.

We conclude that most studies on inter-individual differences in VO_2max response to exercise training have no comparator sample. Therefore, true inter-individual differences in response cannot be quantified, let alone appraised for clinical relevance. For those studies with a comparator sample, we found that the inter-individual differences in training response were not larger than random within-subjects variation in VO_2max over the same time period as the training intervention.

Key Points

Researchers often focus upon ‘main effects’ and mean group changes, but these statistics may hide a wide range of responses.

True inter-individual differences in exercise training response can be precisely quantified and appraised for clinical importance only with parallel information from a suitable comparator group or data from a relevant reliability study. Importantly, none of the studies resulting from the HERITAGE Family Study included a comparator sample.

We provide a ‘road-map’ for the study of inter-individual differences in the responses to exercise training.

1. Introduction

Interest in the concept of individualised responses to an intervention as part of 'personalised medicine' and 'precision care' has been growing over the last 30 years (1,2,3,4,5,6,7,8,9,10,11). In pharmacogenetics, there has been particular interest in 'tailor-made' drugs and therapies, based on the individual response of a patient and/or certain moderators and mediators of that response (8,12). Personalised medicine has also been considered in the context of inter-individual differences in the response of health outcomes to various exercise interventions.

It has been highlighted that the majority of researchers focus upon 'main effects' and mean group changes (13). These statistics are useful, but do not allow us to distinguish between cases (14), may hide a wide range of responses (15) and have previously been described as misleading (13,16). True inter-individual differences in the response to an intervention are less frequently reported, even though it has been proposed that there is large inter-individual variability in response to physical activity interventions (1,2,3,4,5,6,13,17).

Importantly, even in the studies in which inter-individual differences in the response to exercise training are considered, concerns have been levelled at the designs and analysis approaches in these studies (18,19). Therefore, it is important at this time for the claims of inter-individual differences in response to an exercise intervention, with a particular focus on maximal oxygen uptake (VO_2max), to be scrutinised in the context of these recent criticisms. Consequently, we have undertaken the present critical review on the HEalth, RIsk factors, exercise Training And GEnetics (HERITAGE) Family Study, as well as the studies that preceded it and the more recently published research. We focus especially on any apparent limitations of previously adopted data analysis approaches, and how researchers have investigated potential moderators and mediators of the inter-individual difference in VO_2max response to an exercise intervention. Finally, we present what we consider to be an appropriate trial design and analysis approach in order to quantify true inter-individual differences in VO_2max response to exercise interventions. Our focus in this regard is on parallel group randomised controlled trials, as we believe that this design is more widely applicable to research questions addressing chronic adaptations to training. Moreover, published chronic training studies with VO_2max as the outcome are exclusively before-and-after designs, either with or without a control group, with a single intervention period. However, we acknowledge that other designs and statistical approaches have been proposed for quantifying individual differences in response to treatments, primarily the multiperiod (replicate) crossover design (20,21).

2. Maximal oxygen uptake and precision medicine

Low cardiorespiratory fitness has been established as an independent predictor of all-cause mortality and cardiovascular disease (22,23). Many researchers have highlighted the favourable changes in risk factors that occur following a period of exercise training (24,25,26,27). Given that one metabolic equivalent (MET) is the amount of oxygen consumed whilst sitting at rest, and is ≈3.5 ml/kg/min (28), research that 1-MET increase in cardiorespiratory fitness translates to a 12% reduction in cardiovascular disease and all-cause mortality risk has been reported (24).

While a multitude of phenotypes have been investigated, VO_2max response has often been the focus for authors observing the inter-individual variation in response to exercise. Wide inter-individual differences in the trainability of the cardiorespiratory system have been claimed (3,29,30), with reports that the improvements in VO_2max range from zero to a 40% increase (29). Such variation is consistent with the fact that biochemical and

physiological functions vary in all humans (31). Several researchers have also reported that some individuals show little or no improvement in markers such as lipolytic activity, insulin sensitivity, maximal work rate, submaximal exercise heart rate and respiratory exchange rate following an exercise intervention (2,3,4,5,6). Conversely, it has been proposed that physical activity may increase cardiovascular risk in some individuals, worsening risk factors beyond measurement error and biological variation (32), although this notion is not consistent with the results of a more recent study (33).

Prescription of exercise is often undertaken with a global approach rather than a personalised one, and as exercise interventions are often utilized to reduce or prevent age-related reduction in function or lifestyle related diseases, attention should be paid to the response of each participant within a study (34). If an individual is likely to respond favourably to a given stimulus, he/she may more likely to engage with that mode of exercise. Consequently, identifying individuals likely to gain greatest benefit would allow practitioners to also focus on alternative exercise, dietary or pharmacological options for those that may be less likely to respond (35,36).

Insert Table 1 here.

3. A critical review of the relevant studies

Via a search of the relevant literature databases, we aimed to locate all the studies in which inter-individual differences in the response of VO_2max to an exercise intervention have been considered. We were particularly interested in ascertaining how many of these studies incorporated a relevant comparator sample into their design. Data from this sample have been deemed to be important for precise quantification and interpretation of inter-individual differences in response (18,19). Without these data, measurement error and random or biological variation in the study outcome over time can compromise the quantification of true inter-individual differences in response (33). Importantly, any physiological outcome can show substantial natural variability over a 4-6 month follow-up period in a control sample that does not receive the intervention (33). This variation will also be present in the intervention group, irrespective of the additional influence of the intervention itself.

3.1. Pre-HERITAGE Studies

The seminal studies on this topic were conducted in the 1980s, with the aim of identifying the inter-individual response to exercise and to clarify the genotype dependency of the modulation of response (1-6) (Table 1). The effects of a 20-week endurance training programme on maximal aerobic power (MAP), ventilatory aerobic threshold and ventilatory anaerobic threshold in ten pairs of monozygotic twins were initially investigated (1). Unlike in later studies, a comparator (no-exercise training) group was included in this study. From the intraclass correlations (ICC) reported, the authors described a highly variable response to training and concluded that sensitivity to training is genotype-dependent. The authors estimated that 20-25% of training-induced variation in MAP was due to within-pair differences. Nevertheless, using the approach recently described (19), re-analysis of the data presented in Table 1 of this study revealed no clinically important differences in the SD of the change scores between the groups (control ± 5.6 ml/kg/min, intervention ± 3.7 ml/kg/min). This observation indicates that there are no substantial inter-individual differences in response to the intervention (19). In fact, these SDs indicate greater variability in response in the control group versus the intervention group. We have argued previously that this phenomenon may be due to imprecision in the estimation of inter-individual

responses with inadequate sample sizes and/or caused by the intervention having a 'homogenizing' effect on the outcome variable, thus reducing the SD of the changes relative to the control group (19).
Further research was undertaken (2,3,4,5,6,37), with the study authors claiming there to be large variations in response to exercise for a number of phenotypes, including adipose tissue, fat cell weight, lipolytic activity, glucose conversion into fat cell, triglycerides, skinfolds, percentage body fat, anaerobic, alactic and lactic acid capabilities, fibre type, enzyme activity, sensitivity of muscle characteristics and aerobic endurance performance. Crucially, no comparator group was included in these studies.

A variation in improvement in maximal aerobic performance (VO_2peak) of between 5 and 88% that was not correlated with a similarly wide range of 16-97% increases in total work output accomplished in a 90-minute ergocycle performance test was reported in one study (3). Inter-individual responses were concluded following the observation of greater between-pair variation than within-pair variation in monozygotic twins, and through sex differences in those studies using mixed-sex cohorts (2). Genotype dependent responses for both maximal aerobic power and endurance performance were observed in conjunction with skeletal muscle enzyme changes following a fifteen-week training programme (5), while inter-individual differences in anaerobic alactacid (ALC) and lactacid (AAC) response, fibre type changes and enzyme activity were reported in response to high intensity intermittent training (6). ALC and enzyme activity were said to be determined by genotype, although no such relationship was observed for other measured variables. The use of siblings was used to make inferences about the importance of genetic influence in heritability, with *F*-ratios suggesting 5-10 times more variance between twin pairs than within pairs. Similarly, genetic determination has been claimed for several different aerobic performance measures from the results of studies in which brothers, monozygotic and dizygotic twins were compared (37). Changes in aerobic fitness ranging from 0-58% were later reported among adults aged 60-71, where a trend for older participants improving less than younger subjects was observed (38).

The justification for the lack of a non-exercising control in the subsequent HERITAGE Family Study, which appears to have been continued through subsequent investigations, was an observation of mean values from previously studied control groups remaining unchanged (39). However, a finding of no substantial change in the mean for the control group can occur in the face of substantial random within-subject variability in the changes in VO_2max over the duration of the study. The variability in the changes in the intervention group must be assessed against the backdrop of this natural variability. In a randomised controlled trial (RCT), the mean effect of the intervention is given by the mean change in the intervention minus the mean change in the control. This logic should be extended to the assessment of inter-individual responses to an exercise intervention. In a parallel group RCT, one cannot say with 100% certainty whether or not any specific individual in the intervention group is a positive responder, as what would have happened to that person if, contrary to the fact, they had been in the control group is unknown. This is the fundamental counterfactual basis of the RCT. However, if the variance in the response in the intervention group is substantially greater than that in the control arm, then true individual responses may be inferred. The control group variability over the same time period as the intervention effectively provides our best guess of the counterfactual - what would have happened to individuals in the intervention group if they had been in the control arm. In parallel group RCTs, substantially greater response variance in the intervention group versus control is both necessary and sufficient for inferring true inter-individual differences in response to the intervention. Assuming that sample estimates are accurate estimates of

the population values, it is incontrovertible that there must be a larger variance in response in intervention vs. control if true individual differences exist in response to treatment. Furthermore, although a parallel group RCT cannot isolate variance due to subject-by-treatment interaction (21), in this design a greater response variance in intervention vs. control is sufficient to infer inter-individual responses. As described in section 4, for any individual in the intervention arm we can then derive the probability of being a positive responder/ trivial responder/ or negative responder.

3.2. Recent Studies

Six to nine times more variance in VO_2max response between monozygotic twin pairs than within pairs has been reported (40). This and other studies were described as 'standardized and carefully monitored' (9), with a 'careful and constant program of quality control and assurance' (41); yet still lack a suitable comparator sample. Nevertheless, RCTs are not only relevant to the investigation of main effects (20). Use of the intervention-only arm as a basis for analysis is problematic, as similar or even greater variability of changes may also be observed in a control group, as was the case when a previous study was re-examined (1). We fear that too much emphasis has been placed on gene relationship statistics without answering the initial and crucial question of whether clinically-relevant inter-individual differences in response exist. This question is answered by calculating the difference in baseline to follow-up variability between intervention and comparator groups, and comparing this difference to a rationalised minimum clinically important difference (MCID) (19).

More recently, large variations in VO_2max response to exercise were reported in the large scale Dose-Response to Exercise in Women (DREW) Study (42). A decrease in the prevalence of non-response with increased training volume was also observed. The authors reported a large amount of inter-individual variability (-33.2 to 76.0% change), citing baseline VO_2max, age and training volume as predictors of non-response. The study comprised three intervention groups (4, 8 and 12 kcal/kg per week of exercise) alongside a control group, with a stated purpose of the analysis to examine the determinants of change in VO_2max in response to exercise training. However, the decision to exclude the control group from the analysis compromised the correct quantification of the true inter-individual response and missed potentially vital information. Further work from the DREW study reported that 30% of participants experienced no improvement in VO_2peak (43). However, once again, no control group data were studied.

Recent studies have been undertaken to further identify possible genotype or phenotype interactions responsible for moderating the magnitude of inter-individual response (15,17,44). Large variation in training response to an eight-week aerobic endurance training intervention was reported (17). Interestingly, whilst a control group was used, the baseline to follow-up changes in this group were not used for comparison at all. As we have stated, disregarding the control group in this manner, on the basis that there will be no mean change and/or the short-term test-retest reliability is high, is a flawed approach. Differences in response were observed by dividing the intervention group responses by quartile, rather than retaining a continuous variable; this approach discards information and has previously been reported to be an inadequate analysis method in epidemiology (45). Upon closer inspection, in contrast to the authors' assertion of the differential effects of the training protocol on the sympathetic nervous system, it appears that there may be little true difference in the variation (SD_{change}) between each of the 'response' groups.

3.3. Concurrent Training

Investigations into the inter-individual responses to combined endurance and strength training in young (46) and older adults (15) have also been undertaken. The findings of these studies are in general agreement with much of the previously published literature, in that a range of training responses were observed. Nevertheless, as in an earlier study (1), a control group was included in one study (15) but no specific comparison was made. It is also apparent from the responses reported in Figure 1 of this investigation (15) that similar variation in response exists in the control group as in the experimental groups, reinforcing the view that there is similar variability of baseline to follow-up changes across all groups. The participants in another study acted as their own control in a crossover trial (46), however, the residual training effect of the intervention period on the response following the washout period is unknown.

3.4. Biological Variability

Not all inter-individual response may be due to the factors postulated in the studies reviewed within this article. Neither does variation in responses confirm that this assumption of inter-individual difference in response is true for any particular study (19). Small day-to-day changes cannot be classified as a worthwhile change, and the response must be clinically relevant and more than the natural biological variation between baseline and follow-up measurements (47). Of course, patients differ not only by genetics, but also by their personal history and environmental circumstances (48), and this can lead to a multitude of effects on inter-individual response. There appears to be little doubt that the response to exercise training is influenced by multiple factors.

A new focus on the quantification of true inter-individual differences and the moderators and mediators responsible may, therefore, have substantial clinical relevance, with any correctly quantified heterogeneity affording the opportunity to identify possible molecular determinants (49). Indeed, RNA profiling may be a potential methodology for capturing information critical to informing the integrated physiological response and molecular determinants (36), once the presence of inter-individual variation in response has been confirmed.

3.5. Identifying ‘Responders’ and ‘Non-Responders’

A further limitation of much of the previous research is the classification of individuals as ‘non-responders’ (49,50) without first defining the term, although this has been partially addressed more recently when defined as those improving by ‘less than the natural biological variability of the selected variable’ (47). Strictly, a positive response should be defined as an increase that is greater than the MCID. For VO_2max, for example, the MCID could be defined as 1 MET, anchored to a clinically relevant relative risk reduction for all-cause mortality of around 12% for this value (24). For a given individual, the observed change in VO_2max after the intervention can be combined with knowledge of the natural random variation of VO_2max over the same time period (from a control group or similar reliability study) to derive the probability that this individual’s true response is greater than the MCID (18). We can then more properly describe each individual in the intervention as, for example, ‘likely to be a responder’, ‘very unlikely to be a responder’, ‘possibly negative responder’, and so on.

Similarly to previous reports (42), further argument for the dose-response to exercise is observed when greater exercise volume (43) or intensity (44) was associated with reduced chances of being classified as a non-responder. While direct comparison between studies is not straightforward, these and similar findings suggest

that some people may be more sensitive to dose prescription of exercise, as opposed to being non-responsive. If this is the case, effective identification of dose requirement or requirement for multimodal approaches such as concurrent training may provide a capability for enhancing the efficacy of an intervention (51). However, as is covered in this review, we argue that an individual cannot be categorically defined as a 'responder' or other such descriptor; merely a probability (percentage chance) that they are such can be applied to each individual (18). Even with this information, in a single-period before-and-after study design, this process can only occur in the presence of an appropriate comparator group assessed over the same or very similar time period as the exercise intervention.

3.6 The HERITAGE Family Study

The large-scale, longitudinal, multicentre HERITAGE Family Study was initiated to investigate and identify the role of genotype in cardiovascular, metabolic and hormonal responses to a 20-week aerobic exercise training programme (52). The contribution of regular exercise to changes in cardiovascular disease (CVD) and diabetes mellitus risk factors was also investigated (41,52). To date, 186 separate publications have resulted from the study, with some of these involving the comparison of various familial relationships to determine the relative importance of genetics (53,54,55). The bulk of the research undertaken during HERITAGE asserts that there are no genotype-specific covariate effects on VO_2max response, such as age, sex or weight (13,30,50). Familial aggregation was reported in response to maximal (3,49,56) and submaximal (54,57) aerobic training, with two and a half times more variance reported between than within families for the VO_2max response.

A reported genetic contribution of approximately 47% (49,55) of the variance observed in VO_2max training response from the HERITAGE studies was similar to that previously reported (3,58). These data were characterized by a strong maternal aggregation (49,54,56), with shared environmental factors also contributing to the observed heritabilities. The mechanisms underpinning this variance are unclear, but suggestions of genetic contribution from mitochondrial DNA (49) or expression of genes inherited from the mother have been presented (54). Correlations between spouses have led to familial environmental factors also being postulated as responsible for some of the variance observed in response to exercise (54,56,58,59); however the correlations presented are small (r=0.14-0.26), therefore posing, rather than answering, further questions on this issue. The crucial question that is, again, unanswered in the absence of a comparator group is whether there are genetic influences on individual magnitude of random within-subjects variability. If this is known, then these genetic influences could be quantified.

In HERITAGE studies, it is claimed that there is considerable variation in response. Nevertheless, it remains unclear as to whether it was the same individuals that showed no response for all measures, or if each individual showed differing response characteristics across the spectrum of physiological markers investigated. Recent research has attempted to elucidate this issue, observing improvements in at least one measured variable in every individual (47), though again, this study is limited by the lack of a control group with which to compare the inter-individual response. Interestingly, despite methodological concerns, individuals with the highest response to endurance training have also shown high response to resistance training, but the reverse was not true (46). This area opens up future avenues in which to investigate the magnitude of response, in the presence of proper initial quantification through comparison with a suitable comparator sample.

3.7. Twin Studies

Previously noted criticisms of twin studies point to the fact that they may not necessarily separate genetic from environmental pathways (60). Presentation of evidence of genetic variance of numerous phenotypes through greater between-twin pair variance than within-pair variance is often reported through the use of ICCs (1,5,36,61,62,63,64,65,66) (Table 2). These observations are highly sample-specific, and a comparison of the ICC between studies is not without difficulty, due to the heterogeneity of samples. The potential for ICCs in twin studies to overestimate heritability has also been highlighted in that genetic and environmental factors have not been adequately separated (63).

It is as yet unknown as to whether any relationship between genes and phenotype is even linear (60) and while genetic variation is not denied in this review, it is not clear that all observed variance is genetic (48). We believe that when analysing such a design, it would be more appropriate to use data from a relevant control (no-exercise) sample and a linear mixed model in order to correctly quantify the influence of genetics on magnitude of response. Associations could then be presented as a regression coefficient in the units of measurement, rather than a comparison of correlation values. In this way, the clinical importance of any association can be inferred.

Common underlying environmental effects have also been proposed as being underestimated due to study design or low statistical power (67). Adoption studies combined with twin studies to compare identical and fraternal twins and twins reared apart (63) and repeated assessments (20) may be required to quantify some of these issues.

Insert Table 2 here.

3.8. Baseline Correlation of Changes

Several authors of HERITAGE studies correlated each individual's baseline score with the follow-up change to attempt to determine the contribution of baseline status to the inter-individual response to exercise training. From such analyses, it has also been reported that age, sex, fat mass, fat-free mass, weight and race have little or no impact upon the inter-individual response to training or covariate effect (13,30), and that the initial level of the phenotype was a major determinant of the magnitude of response in some cases (13). Nevertheless, this correlation approach has been questioned, due to regression to the mean and mathematical coupling influences (68). Linear mixed effects modelling and other methods such as that previously reported (69) have been purported to be superior to this simple correlation approach (68).

3.9. Testing Quality Control

The HERITAGE intervention was described as having a careful and constant program of quality control and quality assurance (41). Nevertheless, this claim was based on the test-retest mean differences being small, although the selection of either an average of two VO_2max test scores (where coefficient of variation (CV) was less than 5% between the two) or the higher score (if CV was greater than 5% between the two) at both baseline and follow-up (70) could have led to inconsistent data. To accurately analyse the data, identical methodology should ideally be used for all participants. Test-retest reliability was reported to be 4.1-5.0% and ICCs of 0.96 to 0.97 were reported over a period of two days (70) and two weeks (71), implying adequate short-term reproducibility. It is our belief that reproducibility needs to be assessed over a longer period, preferably

matching the length of the intervention, in order to estimate the true extent of longer-term within-subject variation. A better alternative is to use an RCT design, wherein the control group in effect acts as the perfect contemporaneous reliability study. Each of the investigations discussed have contained a single application of an intervention (single period before-and-after study). We reiterate that the primary limitation of the parallel group RCT design in permitting the quantification of inter-individual variation in treatment response is that it does not allow the isolation of the variance due to true subject-by-treatment interaction (21). In this design, the SD for inter-individual responses – although free from random error - includes the subject-by-treatment interaction plus any within-subject variability in treatment response introduced by the intervention (20). Indeed, the multiperiod (replicate) crossover study, in which participants are randomised to sequences in which they receive both the intervention and comparator treatments in at least two periods each, is the only design that can identify variance between-treatments, between-subjects, and the subject-by-treatment interaction (20). However, the primary limitation of the replicate crossover, in the context of chronic training studies, is the long and uncertain washout periods required and hence potentially substantial carryover effects (18).

The authors of a recent investigation into the cardiac determinants of individual response in change in aerobic fitness after a moderate intensity exercise intervention (43) stated that they incorporated 'well-controlled exercise trials' in keeping with the HERITAGE study. Nevertheless, 'well-controlled' appears to refer to relatively short-term repeatability of measurements (over a few days) rather than the within-subjects variability in measurements over the duration of the intervention (a few months). Just because a measurement method has good short-term repeatability does not rectify the problem of lack of a control group, which must be employed in order to make a formal comparison of the variability of the change scores in intervention vs. control groups. Consequently, the inclusion of data from studies such as these is potentially misleading, and as such, participants from these studies that have been termed 'responders' and 'non-responders' may have been selected for further investigation as to the potential moderators and mediators of the inter-individual response, when it may be nothing other than their natural biological variation that has been measured.

3.10. *N*-of-1 Trials

In pharmacogenetics, *n*-of-1 trials have been proposed (72,73), but these single-subject trial studies have previously been linked to controversial issues in clinical investigation, such as carryover effects and the presupposition of patient-by-treatment interaction and behavioural change (74) that may confound the effectiveness of interventions. Of course, if a number of *n*-of-1 trials are carried out, then the combined data effectively equates to the repeated period crossover design proposed by Hecksteden et al. (20). It has also been proposed that *n*-of-1 data with a limited observation count per participant may not be compatible with statistical models that aim to identify the inter-individual response and may be preferential for estimating the population effect (75).

4. A Road Map For Future Study Designs and Analyses

Recently, for both parallel group and replicate crossover designs, more appropriate and robust statistical approaches have been forwarded for the quantification of true inter-individual response to a treatment. Relevant sources of variability must first be quantified (20) before any exploration is undertaken of the true inter-

individual variation in treatment response. Additionally, without knowledge of the smallest worthwhile change or the MCID, no substantial inter-individual differences in VO_2max response to an exercise intervention can be claimed. When analysing the collected data from a parallel group RCT, it has been proposed that comparing the standard deviation of the intervention arm of the study against the standard deviation of the comparator arm, using $SD_R = \sqrt{SD_I^2 - SD_C^2}$, where R = inter-individual responses, I= the intervention sample SD and C = the comparator sample SD (18,19), provides a more accurate statistical analysis of the presence of inter-individual differences in response. If appropriate clinical inferences are to be made about the magnitude of change and any inter-individual response to the intervention, standard deviations, confidence intervals, effect sizes and magnitude-based inferences should also be interpreted (18,76). Using a custom spreadsheet (77), and with knowledge of the typical error over the same timeframe as the intervention and the smallest worthwhile change, the probability (percentage chance) of each individual being classified as 'very likely', 'likely', 'possibly', 'possibly not', 'unlikely' and 'very unlikely' to be a responder can be calculated. This is a more robust approach, as the standard parallel arm study design renders the definitive identification of specific individuals as non-responders impossible (33). For instance, individuals could be termed likely 'positive' responders if the individual probabilities were above 0.75 (75% chance, or odds of 3:1 in favour) and the converse for 'negative' responders. A finding of substantial clinically relevant inter-individual differences in response to the intervention would justify further investigation of potential moderators and mediators, using more advanced statistical modelling.

If we consider the original pre-HERITAGE study (1), the mean VO_2max improvement in the exercise intervention group was 5.5 (± 3.7) ml/kg/min and the change in the control group was -0.6 (± 5.6) ml/kg/min. The pooled between-subjects SD for VO_2max at baseline was 5.9 ml/kg/min. If we define a 'responder' by an improvement of 1 MET, an individual would be required to improve by 7.4 ml/kg/min (i.e. approximately 1.25 SDs) for the probability of being a true responder to be 0.75. To increase confidence, using a probability of 0.95 (i.e. 'very likely' to be a responder), the individual would be required to improve by 13.5 ml/kg/min, or more than 2 SDs. Therefore, an individual who showed an improvement of, say, 5 ml/kg/min (a figure above the clinically relevant threshold for a responder of 1 MET) would have a probability of 0.60 of being a true responder. Obviously, in this case, this is little better than chance. These figures demonstrate that an individual would be required to improve their VO_2max substantially more than the MCID (i.e. 1 MET) in order to be deemed likely or very likely to be a responder. This is in stark contrast to the practice of classification of any individual showing improvement of 3.5 ml/kg/min (1 MET) or more as a definite responder. Assuming normal distribution of the changes in the control group and a MCID of 1 MET, the mean and SD reveal that 23% of the control group would be expected to 'improve' by more than 1 MET, and would be labelled conventionally as 'positive responders'. These apparent positive responses in the control are due to the random variation in VO_2max over a 20-week period. As highlighted, the SD of the change scores in each group reveal that there are no substantial inter-individual responses in the intervention group (vs. control), and any further investigation of the mechanisms underpinning inter-individual response is therefore unwarranted.

In contrast to our proposed approach, it has been argued that a large-scale multiperiod crossover training study approach is a more robust method of predicting training response (20). This approach, however, presents a number of challenges. Given the difficulties of recruiting the sample size required for a large-scale training

study, this type of study is likely to be statistically underpowered, while the time required to run a training intervention study, complete with washout periods, is highly restrictive. The crossover trial methodology might also have less relevance in training studies than in pharmacological research, as the effectiveness of any washout period is unknown, and may diminish training related effects. This approach has been previously utilised through the use of a two-month washout period subsequent to a two-week intervention (46), but the effects of the previous training intervention cannot be controlled for, and therefore each participant is potentially beginning from a different baseline. Unlike in pharmacological studies, where the washout period for specific drugs is defined as some multiple of the drug's half-life, it cannot be stated with any certainty that a previous period of training or an exercise intervention has not changed the individual at the cellular or neuromuscular level. This problem leads to a sample that is not acting as its own control, and therefore presents potential differences at baseline for each intervention period. The multiperiod crossover design might be more applicable to the investigation of acute effects of short-term interventions (15). There are also a multitude of sources of variability that create challenges in identifying true inter-individual differences in response, such as maturation, diet modulation, disease, lifestyle and environment to be accounted for, further confounding the issue (51).

5. Conclusions

To date, the investigation of inter-individual differences in VO_2max response to exercise training has been conducted almost exclusively without a control group or comparator arm. While we do not deny that the identification of any inter-individual response to an exercise intervention is important, we maintain that the variation must be appropriately quantified prior to deeper investigation, and we recognize that a number of challenges exist in realising this goal. Primary among these is the proper quantification and determination of a threshold for meaningful magnitude of change, to establish the presence of clinically important differences in response (51). In order to quantify the inter-individual response to an exercise intervention, studies should contain the presence of a comparator arm, preferably as an RCT design. A number of variables and health outcomes should also be collected, as some participants may improve across some but not all physiological measures. Furthermore, the correct statistical analysis and modelling must be used in order to identify the presence of true, clinically relevant, individual response, as unless true inter-individual response exists, it is futile looking for treatment interactions (14).

Future work on any primary outcome in exercise intervention trials should focus upon a thorough systematic review of the available literature, in order to determine the robustness of the published data addressing inter-individual differences in response to exercise training. Secondary analysis of the data presented by fellow researchers should also be undertaken, in order to quantify inter-individual responses in previous trials. Only when these effects have been properly quantified, using the standard deviation of the change score (SD_{change}) after adjusting for random within-subjects variability using the following equation: $SD_R = \sqrt{SD_I^2 - SD_C^2}$ (18,19) can the design of experiments to further elucidate the mechanisms responsible for the individual response be confirmed. Supplementary investigations and robust data analysis must then be carried out, using a logical framework (Fig. 1) such as that previously proposed (19) in order to properly identify whether specific moderators and mediators exist that control the likelihood of an individual responding to an exercise intervention, rather than looking to unravel complex gene responses. At this point, when included as covariates,

these moderators and mediators may account for the inter-individual response, to the extent that they reduce the magnitude of the SD for inter-individual responses (15).

Insert Fig. 1 here

In summary, against the backdrop of suggestions of precision interventions, individuals may respond to treatment in a variety of ways; the intervention might be beneficial, ineffective, or harmful for different people. The issue of inter-individual differences in the response of maximal oxygen uptake following an exercise intervention is very important, and identifying the personal characteristics that account for these variations in response may ultimately allow more effective direction of interventions. Common themes in previous trial design and data analysis are evident, such as a lack of comparator arm or disregarding data from the control, and the use of ICCs to quantify genotype dependency of inter-individual difference in the variability of VO_2max response. While the subject is an important one, it is crucial that the correct quantification methodology is employed, together with an understanding of the clinical importance of any inter-individual response, before suggestions can be made in regard to potential moderators and mediators responsible for the observed inter-individual variance of VO_2max in response to exercise training.

Compliance and Ethical Standards

Funding

No sources of funding were used in the preparation of this article.

Conflict of Interest

Philip Williamson, Greg Atkinson and Alan Batterham declare that they have no conflicts of interest that are relevant to the content of this review.

References

1. Prud'Homme D, Bouchard C, LeBlanc C, et al. Sensitivity of maximal aerobic power to training is genotype-dependent. Med Sci Sports Exerc. 1984;16:89-93.

2. Despres JP, Bouchard C, Savard R, et al. The effect of a 20-week endurance training program on adipose-tissue morphology and lipolysis in men and women. Metabolism. 1984;33:235-9.

3. Lortie G, Simoneau JA, Hamel P, et al. Responses of maximal aerobic power and capacity to aerobic training. Int J Sports Med. 1984;5:232-6.

4. Savard R, Despres JP, Marcotte M, et al. Endurance training and glucose conversion into triglycerides in human fat cells. J Appl Physiol (1985). 1985;58:230-5.

5. Hamel P, Simoneau J, Lortie G, et al. Heredity and muscle adaptation to endurance training. Med Sci Sports Exerc. 1986;18:690-6.

6. Simoneau JA, Lortie G, Boulay MR, et al. Inheritance of human skeletal muscle and anaerobic capacity adaptation to high-intensity intermittent training. Int J Sports Med. 1986;7:167-71.

7. Rose EA, Partfitt G. A quantitative and qualitative explanation of the individual differences in affective responses to prescribed and self-selected exercise intensities. J Sport Exerc Psychol. 2007;29:281-309.

8. Senn S, Rolfe K, Julious SA. Investigating variability in patient response to treatment – a case study from a replicate cross-over study. Stat Methods Med Res. 2011;20:657-66.

9. Bouchard C. Genomic predictors of trainability. Exp Physiol. 2012;97:347-52.

10. Bouchard C, Antunbes-Correa LM, Ashley EA, et al. Personalised preventive medicine: genetics and the response to regular exercise in preventive interventions. Prog Cardiovasc Dis. 2015;57:337-46.

11. Mann TN, Lamberts RP, Lambert MI. High responders and low responders: factors associated with individual variation in response to standardized training. Sports Med. 2014;44:1113-24.

12. Spear BB, Heath-Chiozzi M, Huff J. Clinical applications of pharmacogenetics. Trends Mol Med. 2001;7:201-4.

13. Bouchard C, Rankinen T. Individual differences in response to regular physical activity. Med Sci Sports Exerc. 2001;33,S6:S452-3.

14. Senn S. Individual response to treatment: is it a valid assumption? BMJ. 2004;329:966-8.

15. Karavirta L, Hakkinen K, Kauhanen A, et al. Individual responses to combined endurance and strength training in older adults. Med Sci Sports Exerc. 2011;43:484-90.

16. Bouchard C. Human adaptability may have a genetic basis. In: Landry F, editor. Health and risk estimation, risk reduction and health promotion. Proceedings of the 18th annual meeting of the society of prospective medicine. Ottawa: Canadian Public Health Association; 1983. pp.463-476.

17. Hautala AJ, Makikallio TH, Kiviniemi A, et al. Cardiovascular autonomic function correlates with the response to aerobic training in healthy subjects. Am J Physiol Heart Circ Physiol. 2003;285:1747-52.

18. Hopkins W. Individual responses made easy. J Appl Physiol (1985). 2015;118:1444-6.

19. Atkinson G, Batterham A. True and false interindividual differences in the physiological response to an intervention. Exp Physiol. 2015;100:577-88.

20. Hecksteden A, Krauscher J, Scharhag - Rosenberger F, et al. Individual response to exercise training – a statistical perspective. J Appl Physiol (1985). 2015;118:1450-9.

21. Senn S. Mastering variation: variance components and personalised medicine. Statist. Med. 2016;35:966–77.

22. Laukkanen JA, Kurl S, Salonen R, et al. The predictive value of cardiorespiratory fitness for cardiovascular events in men with various risk profiles: a prospective, population based cohort study. Eur Heart J. 2004;25:1428-37.

23. Sui X, LaMonte MJ, Blair SN. Cardiorespiratory fitness as a quantitative predictor of non-fatal cardiovascular events in asymptomatic men and women. Am J Epidemiol. 2007;165:413-23.

24. Myers J, Prakash M, Froelicher V, et al. Exercise capacity and mortality among men referred for exercise testing. N Engl J Med. 2002;346:793-801.

25. Church TS, Earnest CP, Skinner JS, et al. Effects of different doses of activity on cardiorespiratory fitness among sedentary, overweight or obese postmenopausal women with elevated blood pressure: a randomized control trial. JAMA. 2007;297:2081-91.

26. Kelley GA, Kelley KS. Efficacy of aerobic exercise on coronary heart disease risk factors. Prev Cardiol. 2008;11:71-5.

27. Church TS, Blair SN, Cocreham S, et al. Effects of aerobic and resistance training on hemoglobin A1c levels in patients with type 2 diabetes: a randomized control trial. JAMA. 2010;304:2253-62.

28. Jette M, Sidney K, Blumchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol. 1990;13:555-65.

29. Bouchard C. Individual differences in the response to regular exercise. Int J Obes Relat Metab Disord. 1995;19,S4:S5-S8.

30. Feitosa MF, Gaskill SE, Rice T, et al. Major gene effects on exercise ventilatory threshold: the HERITAGE family study. J Appl Physiol (1985). 2002;93:1000-6.

31. Vollaard NBJ, Constantin-Teodosiu D, Fredriksson K, et al. Systematic analysis of adaptations in aerobic capacity and submaximal energy metabolism provides a unique insight Into determinants of human aerobic performance. J Appl Physiol (1985). 2009;106:1479-86.

32. Bouchard C, Blair CN, Church TS, et al. Adverse response to regular exercise: is it a rare or common occurrence? Plos One. 2012;7.

33. Leifer ES, Mikus CR, Karavirta L, et al. Adverse cardiovascular response to aerobic exercise training: is this a concern? Med Sci Sports Exerc. 2015;48:20-5.

34. Kainulainen H. Run more, perform better – old truth revisited. J Appl Physiol (1985). 2009;106:1477-8.

35. Rankinen T, Argyropoulos G, Rice T, et al. CREB1 is a strong genetic predictor of the variation in exercise heart rate response to regular exercise: the HERITAGE family study. Circ Cardiovasc Genet. 2010;3:294-9.

36. Timmons JA, Knudsen S, Rankinen T, et al. Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. J Appl Physiol (1985). 2010;108:1487-96.

37. Bouchard C, Lesange R, Lortie G, et al. Aerobic performance in brothers, dizygotic and monozygotic twins. Med Sci Sports Exerc. 1986;18:639-45.

38. Kohrt WM, Malley MT, Coggan AR, et al. Effects of gender, age, and fitness level on response of VO_2max to training in 60-71 yr-olds. J Appl Physiol (1985). 1991;71:2004-11.

39. Wilmore JH, Stanforth PR, Gagnon J, et al. Heart rate and blood pressure changes with endurance training: the HERITAGE family study. Med Sci Sports Exerc. 2000;33:107-16.

40. Bouchard C, Wolfarth B, Rivera MA, et al. Genetic determinants of endurance performance. In: Shephard RJ, Astrand PO, editors. Endurance in sport: encyclopaedia of sports medicine. Oxford: Blackwell Scientific; 2000. pp.223-42.

41. Gagnon J, Province MA, Bouchard C, et al. The HERITAGE family study: quality assurance and quality control. Ann Epidemiol. 1996;6:520-9.

42. Sisson SB, Katzmaryk PT, Earnest CP, et al. Volume of exercise and fitness non-response in sedentary post-menopausal women. Med Sci Sports Exerc. 2009;41:539-45.

43. Pandey A, Ayers C, Blair SN, et al. Cardiac determinants of heterogeneity in fitness change in response to moderate intensity aerobic exercise training. J Am Coll Cardiol. 2015;65:1057-8.

44. Ross R, de Lannoy L, Stotz PJ. Separate effects of intensity and amount of exercise on interindividual cardiorespiratory fitness response. Mayo Clin Proc. 2015;90:1506-14.

45. Bennette C, Vickers A. Against quartiles: Catergorization of continuous variables in epidemiologic research, and its discounts. BMC. 2012;12:1. DOI: 10.1186/1471-2288-12-21.

46. Hautala AJ, Kiviniemi AM, Makikallio TH, et al. Individual differences in the responses to endurance and resistance training. Eur J Appl Physiol. 2006;96:535-42.

47. Scharhag-Rosenberger F, Walitzek S, Kindermann W, et al. Differences in adaptations to 1 year of aerobic endurance training: individual patterns of nonresponse. Scand Med Sci Sports. 2012;22:113-8.

48. Senn S. Individual therapy: new dawn or false dawn? Drug Info J. 2001;35:1479-94.

49. Bouchard C, An P, Rice T, et al. Familial aggregation of VO_2max response to exercise training: results from the HERITAGE study. J Appl Physiol (1985). 1999;87:1003-8.

50. Skinner JS, Jaskolski A, Jaskolska A, et al. Age, sex, race, initial fitness and response to training: the HERITAGE family study. J Appl Physiol (1985). 2001;90:1770-6

51. Buford TW, Roberts MD, Church TS. Toward exercise as personalised medicine. Sports Med. 2013;43:157-65.

52. Bouchard C, Leon AS, Rao DC, et al. The HERITAGE family study: aims, design and measurement protocol. Med Sci Sports Exerc. 1995;27:721-9.

53. Bouchard C, Rankinen T, Chagnon YC, et al. Genomic scan for maximal oxygen uptake and its response to training in the HERITAGE family study. J Appl Physiol (1985). 2000;88:551-9.

54. Perusse L, Gagnon J, Province MA, et al. Familial aggregation of submaximal aerobic performance in the HERITAGE family study. Med Sci Sports Exerc. 2001;33:597-604.

55. Bouchard C, Sarzynski MA, Rice TK et al. Genomic predictors of the maximal O_2 uptake response to standardized exercise training programs. J Appl Physiol. 2011;110:1160-70.

56. Bouchard C, Daw E, Warwick E, et al. Familial resemblance for VO_2max in the sedentary state: the HERITAGE family study. Med Sci Sports Exerc. 1998;30:252-8.

57. Gaskill SE, Rice T, Bouchard C, et al. Familial resemblance in ventilatory threshold: the HERITAGE family study. Med Sci Sports Exerc. 2001;233:1832-40.

58. Montoye HJ, Gayle R. Familial relationships in maximal oxygen uptake. Hum Biol. 1978;50:241-9.

59. Lesage RJA, Simoneau J, Jobin J, et al. Familial resemblance in maximal heart rate, blood lactate and aerobic power. Hum Hered. 1985;35:182-9.

60. Maes HHM, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. Behav Genet. 1997;27:325-51.

61. Poehlman ET, Tremblay A, Despres JP, et al. Genotype-controlled changes in body composition and fat morphology following overfeeding in twins. Am J Clin Nutr. 1986;43:723-31.

62. Bouchard C, Tremblay A, Despres JP, et al. The response to long-term overfeeding in identical twins. N Engl J Med. 1990;322:1477-82.

63. Heller DA, De Faire U, Pederson NL, et al. Genetic and environmental influences on serum lipid levels in twins. N Engl J Med. 1993;328:1150-6.

64. Bouchard C, Tremblay A, Despres JP, et al. The response to exercise with constant energy intake in identical twins. Obes Res. 1994;2:400-10.

65. Hong Y, Pederson NL, Brismar K, et al. Genetic and environmental architecture of the features of the insulin-resistance syndrome. Am J Hum Genet. 1997;60:143-52.

66. Tremblay A, Poehlman ET, Despres JP, et al. Endurance training with constant energy intake in identical twins: changes over time in energy expenditure and related hormones. Metabolism. 1997;46:499-503.

67. Segal NL, Allison DB. Twins and virtual twins: bases of relative body weight revisited. Int J Obes Metab Disord. 2002;26:437-41.

68. Chiolero A, Paradis G, Rich B, et al. Assessing the relationship between the baseline value of a continuous variable and subsequent change over time. Front Public Health. 2013;1:1-8.

69. Blomqvist N, Svardsudd K. A new method for investigating the relation between change and initial value in longitudinal blood pressure data. II. comparison with other methods. Scand J Soc Med. 1978;6:125-9.

70. Shephard RJ, Rankinen T, Bouchard C. Test-retest errors and the apparent heterogeneity of training responses. Eur J Appl Physiol. 2004;91:199-203.

71. Skinner JS, Wilmore KM, Jaskolska A, et al. Reproducibility of maximal exercise test data in the HERITAGE family study. Med Sci Sports Exerc. 1999;31:1623-8.9

72. Guyatt GH, Heyting A, Jaeschke R, et al. N–of-1 randomized trials for investigating new drugs. Control Clin Trials. 1990;11:88-100.

73. Lillie EO, Patay B, Diamant J, et al. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? Per Med. 2011;8:161-73.

74. Senn S. Suspended judgement of n-of-1 trials. Control Clin Trials. 1993;14:1-5.

75. Zucker DR, Ruthazer R, Schmid CH. Individual (n-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. J Clin Epidemiol. 2010;63:1312-23.

76. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. Int J Physiol Perform. 2006;1:50-7.

77. Hopkins WH. Precision of the estimate of a subject's true value (Excel spreadsheet). In: a new view of statistics. www.sportsci.org. Internet society for sport science. 2000. sportsci.org/resource/stats/xprecisionsubject.xls. Accessed 21 Nov 2016.

Table. 1. Early studies presenting inter-individual response to exercise interventions.

		Exercise training program			Results	
Study	**Subjects/groups**	**Mode**	**Length**	**Intensity/frequency/duration/volume**	**Δ BW/VO_2max/lipids**	**Other**
Prud'Homme *et al.,* 1984 (1)	*n* = 48 (10 pr (6M, 4F) MZ twins & 14 (7M, 7F) control)	Cycling	20 wk	4-5d/wk; 40-45mins; 60-85% HRR	Variable response, claims that sensitivity to training is genotype dependent	20-25% of training induced variation in MAP due to within-pair differences
Despres *et al*., 1984 (2)	*n* = 22 (11M, 11F)	Cycling	20 wk	4-5d/wk; 40mins; 80% MHR	No Δ in fat cell number. Δ fat cell weight. Δ lipolysis	Δ lipolysis response greater in males than females. Females had no Δ in fat mass, skinfolds. Increased MAP.
Lortie *et al*., 1984 (3)	*n* = 24 (13F, 11M)	Cycling	20 wk	4-5d/wk; 40-45mins; 60-85% HRR	Δ MAP/kg 33%; Δ MAC/kg by 51%; Males Δ in MAC/kg 50% more than females.	Range of variation: 5-88% Δ MAP/kg & 16-97% Δ MAC/kg.
Savard *et al*., 1985 (4)	*n* = 24 (13F, 11M)	Cycling	20 wk	4-5d/wk; 40-45mins; 60-85% HRR	Δ Insulin stimulated glucose conversion to triglycerides. Δ in males, but not females. Similar Δ in MAP.	Suggests Δ in modification of fat cell glucose metabolism.
Hamel *et al*., 1986 (5)	*n* = 12 (6 prs MZ twins)	Cycling	15 wks	3-5d/wk; 30-45mins; 60-85% HRR including 1/wk HIIT; 3x10mins; 80-85% with 5mins recovery	Δ in aerobic enzyme activity in wks 8-15. 5-11 x more variation between than within pairs.	No fibre type Δ
Simoneau *et al*., 1986 (6)	*n* = 28 (14 pr MZ twins, (7M pr, 7F pr))	Cycling	15 wk	4-5d/wk; HIIT 10 x 15-30s & 4-5 x 60-90s;HR recovery to 120-130b.min efforts.	Δ T1 fibres, AAC, ALC, enzyme activity & T2 fibres.	Large inter-individual differences, but similar within twins. Genotype suggested as responsible for responsiveness to HIIT on several variables. 65% of ALC associated with genotype. Δ oxidation following HIIT. Fibre type changes independent of genotype.

Pr – pair, M – male, F- female, MZ – monozygotic, HRR – heart rate reserve, MHR – maximal heart rate, HR – heart rate, b.min – beats per minute, HIIT – high intensity interval training, BW – body weight, VO_2max – maximal oxygen uptake, MAP – maximal aerobic power, AAC- anaerobic alactic acid, ALC – anaerobic lactacid, SD – standard deviation.

Table. 2. Twin studies presenting intraclass correlations in analysis of inter-individual response to exercise interventions

ICC – intraclass correlation, SD – standard deviation, MZ – monozygotic, DZ – dizygotic, M – male, F- female, MZA – monozygotic twins reared apart, MZT – monozygotic twins reared together, DZA - dizygotic

Study	Number of twin pairs	Mean (SD) age (years)	Outcome measures	ICC	Age/sex adjustment
Prud'homme *et al.,* 1984 (1)	10 MZ (6F, 4M)	20 (2.9)	MAP, VAT, VANT	0.24-0.74	Not reported
Hamel *et al*., 1986 (5)	6 MZ (3M, 3F)	21 (4)	VO_2max	0.69	Not reported
Bouchard *et al*., 1986 (36)	53 MZ (mixed sex) 33 DZ (mixed sex) 27 male siblings	16-34 (range)	VO_2max	0.85 0.74 0.55	Yes
Poehlman *et al*., 1986 (61)	6 MZ males	19.2 (2.3)	Body composition, fat morphology, fat mass, skinfolds	0.46 – 0.90	Not reported
Bouchard *et al*., 1990 (62)	12 MZ males	21 (2)	Body composition & fat topography	0.4 – 0.55	Not reported
Heller *et al*., 1993 (63)	46 MZA 7 MZT 100 DZA 89 DZT	52-86 (range)	Lipids	0.22 – 0.79 0.33 – 0.83, -0.06 – 0.47 -0.13 – 0.49	Dichotomous age categories
Bouchard *et al*., 1994 (64)	7 MZ males	21 (0.8)	Body weight, FFM, topography, VO_2max	0.25 - 0.87	No mention – single sex, low SD of age
Hong *et al*., 1997 (65)	289 pr; 45 MZA 64 MZT 95 DZA 85 DZT		Insulin, glucose, lipids, BP	0.28-0.64 0.56-0.6 0.17-0.39 0.15 -0.26	Yes
Tremblay *et al*., 1997 (66)	11pr MZ males	21 (0.8)	RMR, fat loss, weight loss, FFM loss	0.32 – 0.69	Single sex, low SD of age

twins reared apart, DZT - dizygotic twins reared together, pr – pair, MAP – maximal aerobic power, VAT – ventilatory aerobic threshold, VANT – ventilatory anaerobic threshold, VO_2max – maximal oxygen uptake, FFM – fat free mass, BP – blood pressure, RMR – resting metabolic rate.

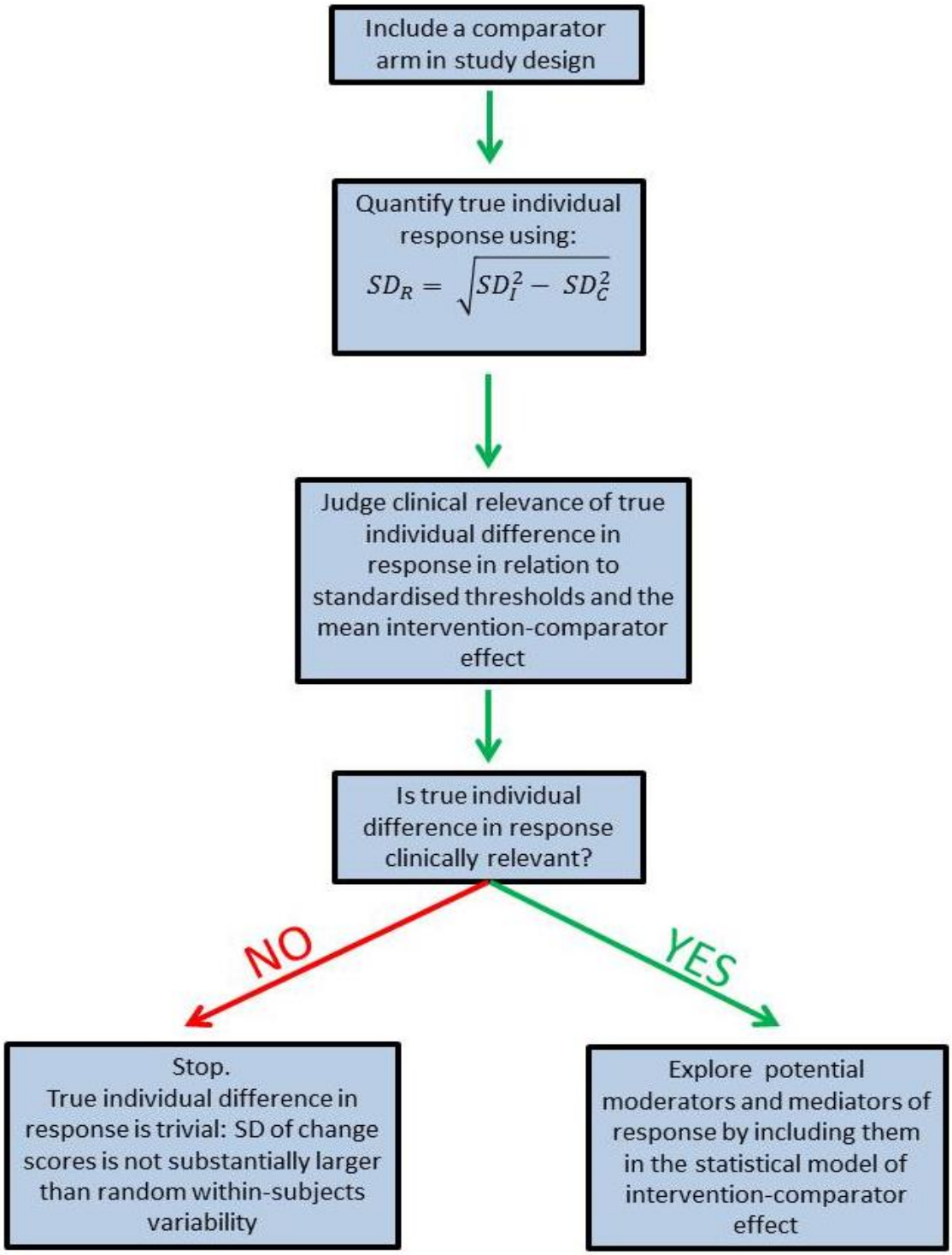

Fig.1. Conceptual framework for the quantification of true inter-individual differences in response to an exercise intervention (19).

SD_R – the standard deviation of the true inter-individual response of the intervention. SD_I – the standard deviation of the pre-to-post change score of the intervention arm. SD_C - the standard deviation of the pre-to-post change score of the comparator arm.