Diabetologia Springer

Exercise training response heterogeneity: Statistical insights. A response to Sparks (2017)

Journal:	*Diabetologia*
Manuscript ID	Diab-17-1627
Manuscript Type:	Letter
Keywords:	3.05.07 Exercise, 3.01 Genetics / Epidemiology (all), 4.11 Biological statistics

Exercise training response heterogeneity: Statistical insights. A response to Sparks (2017)

Greg Atkinson, Philip Williamson and Alan M Batterham

Health and Social Care Institute, School of Health and Social Care,Teesside University, Middlesbrough, UK

All correspondence to:

Greg Atkinson

Health and Social Care Institute

School of Health & Social Care

Constantine Building

Teesside University

MIDDLESBROUGH

Tees Valley TS1 3BA

United Kingdom

Tel: 00-44-1642 342771

Email: g.atkinson@tees.ac.uk

Dear Editor,

We would like to thank Dr. Sparks for her comprehensive review of research on individual differences in the responses to exercise training [1]. It was especially helpful to see some of the statistical issues touched upon in the review [2]. We believe that these statistical factors are crucial for answering the fundamental question of whether there are true and clinically important individual differences in the response to exercise. By "true", we mean individual response differences that are not merely random trial-to-trial variability in disguise. By "clinically important", we mean individual differences that exceed a well-rationalised minimal clinically important difference (MCID). We also maintain that, in the "roadmap" for researching this topic, true and clinically relevant individual response differences should be confirmed empirically *before* any moderators and mediators of the exercise response are explored [3].

The definition of "non-response" was given in [1] as, "*the lack of a difference between a control and a treatment condition with respect to a specific variable*". We are concerned that this definition implies that researchers can identify non-responders simply by looking at their data from a two-condition (control and exercise) experiment and concluding that those participants with a treatment-control difference of zero or less are identified as "non-responders". The fallacy of this approach was hinted at in [2], but the full implications of this issue were not explicitly described. We have provided a full account of the pitfalls in non-responder identification [3, 4], and think that they can complement the useful review by Sparks [1].

An observed response is comprised of the true response as well as random trial-to-trial within-subjects variability [3]. Therefore, observed non-response to exercise, or any other treatment, does not necessarily mean that there has been a true non-response. In Figure 1, we present some simulated data which appears to show that individual participants differed substantially in terms of their acylated ghrelin response to exercise. It is well documented that exercise causes a reduction in the mean concentration of acylated ghrelin [4]. For the simulated data in Figure 1, the mean (SD) reduction in acylated ghrelin was 18.1 (23.1) pg/ml (95%CI: 1.6 to 34.7). Nevertheless, it appears as though there are three non-responders in this sample of 10 participants, according to the definition provided in [1].

In reality, true individual differences in response to exercise do not exist in the data presented in Figure 1. In our simulation we subtracted exactly 25 pg/ml of ghrelin from each participant's control condition measurement. We then added the component of a typical magnitude of random trial-to-trial variability. The trial-to-trial correlation coefficient was 0.77. This random variability in biological measurements from day-to-day or week–to-week is always present and uncontrollable. Importantly, it is this component of variance on its own which makes it look as though individual differences in exercise response exist, when they do not. Naturally, this component can also influence the true mean difference, as it does in our Figure 1 data.

For repeated trial studies, the optimal design for quantifying individual response differences is actually the replicate crossover design [5]. Here, the control and exercise conditions are actually administered at least twice to each participant, with the sequence of the four trials randomised. This design allows the researcher to derive the intervention x participant

interaction term from the statistical model [5], thereby allowing the researcher to isolate the true individual differences in response to exercise. To our knowledge, this design has not yet been used in an exercise context, although we have several such studies on-going at present.

Within-subjects variability in the measured outcome also causes problems when interpreting the results of longer-term exercise training studies [6]. In many of these studies, a control group is either not present or discarded in the data analysis. Plots of individual differences in the baseline-to-follow up change are commonly presented for the exercise training study arm only, as in the top graph of Sparks' Figure 1 [1]. Nevertheless, a very similar graph can usually be plotted for the baseline-to-follow up change in the control group. These data are seldom presented but are crucial for ascertaining whether there are clinically relevant true individual differences in training response [3].

In a parallel group study, true individual differences in exercise response are present only if the standard deviation (SD) of change is substantially larger in the exercise group than the control group. If not, the apparent individual differences in "response" are nothing but baseline-to-follow up within-subjects variability, just as in our Figure 1. This random variability can be large if there are many weeks (>6) between baseline and follow up in the study, and this is common in most studies. We recently also presented a critical review of a selected sample of exercise training studies - depending on whether the relevant data were reported or not, and in which the outcome was the change in VO_{2peak} [6]. We found that very few studies included data from a control group in their analyses. For those studies that

had a control group, we found little evidence that the difference in the SD of changes between intervention and control was clinically important, relative to an MCID of 1 MET.

We maintain that ascertaining whether there are true individual differences in the responses to exercise that are large enough to be clinically relevant is a crucial platform for precision medicine. If the individual differences in response are found to be not clinically important, we question the need to proceed to explore individual moderators and mediators of response. Such explorations could be wasteful in terms of participant time as well as money from a funding body. We have also highlighted in this letter that an understanding of the impact of trial-to-trial or baseline-to-follow up variance is crucial for making robust inferences about individual response differences.

[947 words]

References

1. Sparks LM (2017) Exercise training response heterogeneity: physiological and molecular insights. Diabetologia, Ahead of Press. DOI 10.1007/s00125-017-4461-6
2. Barker RJ, Schofield MR (2008) Classifying individuals as physiological responders using hierarchical modeling. J Appl Physiol 105:555–560
3. Atkinson G, Batterham, AM (2015) True and false inter-individual differences in the physiological response to an intervention. Exp Physiol 100: 577-588.
4. Schubert MM, Sabapathy S, Leveritt M, Desbrow B. (2014) Acute exercise and hormones related to appetite regulation: a meta-analysis. Sports Med 44: 387-403.

5. Senn S, Rolfe K, Julious SA (2011) Investigating variability in patient response to treatment - a case study from a replicate cross-over study. Stat Methods Med Res 20: 657-66.
6. Williamson PJ, Atkinson G, Batterham AM (2017) Inter-Individual Responses of Maximal Oxygen Uptake to Exercise Training: A Critical Review. Sports Med 47: 1501-1513.

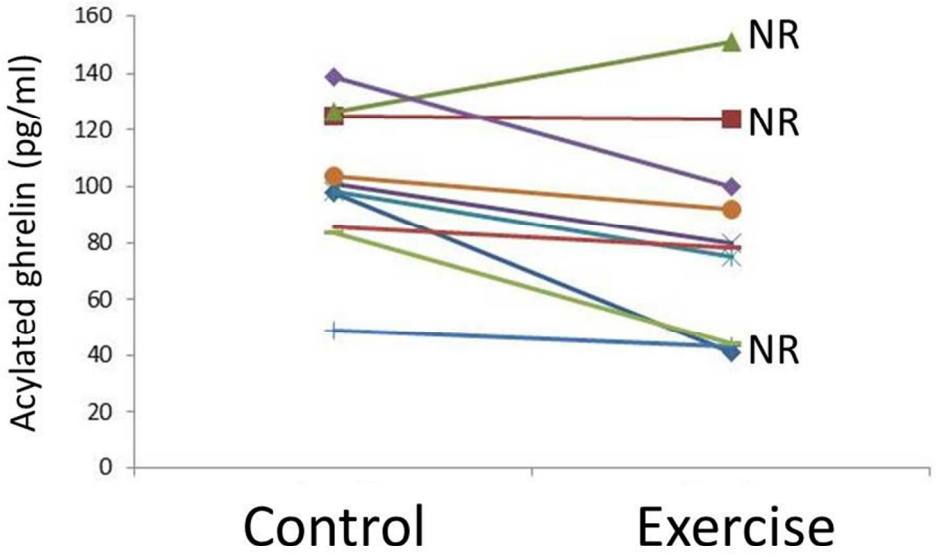

Figure 1. Hypothetical data simulated to show apparent individual differences in the responses of acylated ghrelin concentration to exercise. In reality, there are no true individual response differences at all in these data. Exactly 25 pg/ml of ghrelin concentration was subtract from the control data of every participant. Individual differences only appear to exist in the observed data because of random trial-to-trial variability.

254x190mm (96 x 96 DPI)