

Individual Measurement of Performance Change in Sports

Individuelle Veränderungsmessung im Sport

Summary

- › **Studies** which deal with the change of endurance performance in leisure and health sports, mostly apply the measurements of the mean and dispersion for group comparisons. Also, they interpret significant changes in the performance of an average athlete statistically and practically. However, the average athlete exists only in theory, but not in reality. In accordance with this, individual, often deviating, changes in power and endurance cannot be assessed in a differentiated way, which would be necessary to design an adequate training schedule for each individual athlete. It is also impossible to distinguish accurately between performance change enhancement, biological variability of performance and measurement errors for the individual athlete.
- › **Looking at clinical and training-relevant criteria**, the challenge which is often neglected methodologically is the measurement and evaluation of individual adaptation reactions.
- › **Comparing mean value changes** and a measure for the individual change of a single athlete, the Reliable Change Index (RCI) shows that this may cause training-relevant differences. These differences can be accompanied either by an improvement of individual performance or, in a negative case, by a degradation of individual performance, even if the "average performance" of the group changes just in one direction. The consideration of the individual case is a condition sine qua non condition for effective individual design of training in health- and leisure-oriented endurance training, as well as in other training contexts.

KEY WORDS:

Reliability, Measurement Accuracy, Intraindividual Training Effects, Change Indices, Endurance

Zusammenfassung

- › **Studien**, die sich mit einer interventionsabhängigen Veränderung der Ausdauerleistungsfähigkeit im Freizeit- und Gesundheitssport befassen, beschreiben und vergleichen aufgrund des methodologisch begründeten Mehrgruppenvergleichs Mittelwerte sowie Streuungsmaße und interpretieren statistisch und praktisch bedeutsame Veränderungen der Leistungsfähigkeit eines Durchschnittsaktiven. Der Durchschnittsaktive existiert allerdings nur formal, aber nicht real. Entsprechend können daher auch individuelle, oftmals davon abweichende Veränderungen nicht differenziert beurteilt und adäquate Trainingsprogramme konzipiert werden. Ebenfalls kann für den Einzelnen auch nicht zwischen Leistungsveränderung, biologischer Variabilität der Leistung und Messfehler präzise differenziert werden.
- › **Unter klinisch- und trainingsrelevanten Kriterien** liegt die inhaltlich und methodologisch weitgehend vernachlässigte Herausforderung in der Messung und Beurteilung der individuellen Anpassungsreaktionen, um klinische resp. praktisch bedeutsame Leistungsveränderungen zuverlässig und valide bestimmen zu können.
- › **In einem Vergleich** zwischen Mittelwertsveränderungen und einem individuellen Veränderungsmaß, dem Reliable Change Index (RCI) wird gezeigt, welche trainingsrelevanten Unterschiede resultieren können, die im positiven Fall mit einer individuellen Leistungsverbesserung und im negativen Fall mit einer individuellen Leistungsver schlechterung einhergehen können, obwohl die „mittlere Leistung“ der Gruppe sich nur in eine Richtung verändert. Die Berücksichtigung des Einzelfalls ist eine *conditio sine qua non* für eine effektive individuelle Trainingsgestaltung sowohl im gesundheits- und freizeitorientierten Ausdauertraining als auch in anderen Trainingskontexten.

SCHLÜSSELWÖRTER:

Reliabilität, Messgenauigkeit, intraindividuelle Trainingseffekte, Veränderungsindices, Ausdauer

Introduction

The dynamic planning of training load and the effects on individual change of endurance performance are of great importance not only from the perspective of sports medicine and training science to achieve performance-oriented design of training plans, but also for leisure and health-orientated sports for follow-up and diagnosis of individual load and demand optimi-

zation. In "classical group designs", it is implicitly assumed that individual changes in endurance performance can be represented sufficiently accurately in controlled group designs (see Figure 1, right side, for a simple example) where effect size and confidence intervals for the effect size are calculated in addition to the deductive statistical testing (4). ➤

ACCEPTED: January 2018

PUBLISHED ONLINE: February 2018

DOI: 10.5960/dzsm.2018.315

Ferger K, Büsch D. Individual measurement of performance change in sports. Dtsch Z Sportmed. 2018; 69: 45-52.

1. JUSTUS-LIEBIG UNIVERSITÄT GIESSEN, Institut für Sportwissenschaft, Gießen
2. CARL VON OSSIETZKY UNIVERSITÄT, Institut für Sportwissenschaft, Oldenburg



Article incorporates the Creative Commons Attribution – Non Commercial License. <https://creativecommons.org/licenses/by-nc-sa/4.0/>



QR-Code scannen und Artikel online lesen.

CORRESPONDING ADDRESS:

Dr. Katja Ferger
Justus-Liebig Universität Gießen
Institut für Sportwissenschaft
Kugelberg 62
35394 Gießen
✉: katja.ferger@sport.uni-giessen.de

Table 1

Excerpt from the standard value tables for PWC performance for the age group of 18-35-year-old subjects (27). Age correction for male subjects >30 years old: -1%/year, and for female subjects >30 years: -0.8%/year.

| Norms (watts/kg) PWC150 (18-30 years) | | | | | |
|---------------------------------------|------|------|------|------|------|
| | - | 0 | + | ++ | +++ |
| Men | 1.50 | 2.00 | 2.50 | 3.00 | 3.50 |
| Women | 1.20 | 1.60 | 2.00 | 2.40 | 2.90 |

For example, in a randomized and controlled training study with N=20 participants conducted by the same researchers, it could be demonstrated that a regime of endurance training, comparable between individuals (load: min 1.5 to max. 2.5 Watts/kg body weight; demand: between min. 65% to max. 90% maximum heart rate), led on average to significant improvements (both statistically and in practice) in “physical working capacity” or to the PWC performance at a heart rate of 150 beats per minute ($t[19]=-3.08, p<0.001, d=0.96 [KI_{90\%}:0.2-1.8]$). However, individual scrutiny of the data is advised (see Figure 1, left), both generally also individually due to the relatively large confidence interval for the effect size. For subject 3, deterioration in performance by 0.12 Watts/kg can be seen, whilst the performance change of subjects 1 and 2 is comparable to the group results, with an increase in the individual PWC performances (see Table 1). Subject 1 improves on starting performance by 0.50 Watts/kg, and subject 2 shows an improvement of 0.78 Watts/kg in comparison with starting performance. When interpreting the different improvements, it should be emphasized, however, that both subjects were classified under the same assessment category (i.e. the “+” category) with their final performance scores on the basis of standard PWC₁₅₀ values, despite having different improvement gradients and different starting values (difference between starting performance values of 0.26 Watts/kg) (see Table 1). On the basis of this categorization, an accurate assessment of the differences in the starting performance scores, and particularly of a significant individual performance change between pre-test and post-test scores cannot be considered to be sufficiently reliable.

The calculation of effect sizes for the group comparisons also proves to be unsatisfactory for individual performance changes, given that the effect sizes always also include subjects who showed no improvement or even showed deterioration, and the representation of the average effect can be substantially distorted. Group studies involving larger group sizes elucidate the problem: During a Heritage study, for example, the individual increases in VO₂max varied between 0 and 1000ml/min, but the

effect size (effect size was estimated ex-post from the official data) $d=1.9 (KI_{90\%}:1.8-2.0)$ implied a significant improvement in performance in practical terms for the N=720 participants (2). Contrasting with this, only a trivial training effect of $d=0.4 (KI_{90\%}:-0.01-1.0)$ becomes evident after a 2-week controlled endurance training plan followed by healthy male and female subjects (1), despite 90% of individuals showing at least a moderate improvement in VO₂peak (7). A reason for the lack of adaptation seen in individual subjects could be due to collecting data only on target values. In this regard, measurement of several other parameters pertaining to different adaptation and measuring levels was proposed, as not all parameters react in the same way to endurance training in individual cases. However, this multivariate approach does not solve the problem of different levels of adaptation according to each individual, but rather exacerbates the problem, as the individuality for each parameter, e.g. the heterochronicity of adaptations, as well as time-dependent and effect-dependent interactions would need to be taken into account. For this reason, vis-a-vis empiric “shotgun designs”, parameters selected justified on a theoretical-content basis which can explain the different levels of adaptation according to each individual, seem more promising. Methods for quantification and prediction of individual performance changes, such as methods using single-case diagnosis, offer the opportunity to make differentiated and specific statements for each individual subject regarding the individual point in time, as well as the magnitude of any statistically and practically significant training effect.

General Considerations Regarding Single-Case Diagnosis for Intervention Studies

Alongside designs based on group statistics, single-case diagnoses have played an important role both within clinical effectiveness research and neuropsychology (and indeed in other fields) over the past two decades (5, 12, 20, 32). These methods can be differentiated on the basis of whether they deal with standardized differential values, modified differential values, or regression analysis values (32). All of these methods compare measured values from given points in time (at which measuring was carried out) with one another, and calculate measured values with regards to a variation index. However, the dividend on different calculation rules was relativized for standard error (divisor). A frequently-applied and proven method for assessment of clinically significant changes is the classical method from Jacobson and Truax (13), in which the Reliable Change Index (RCI) is calculated. This concept takes into account (as do all others) that changes cannot be 100% complete, and are calculated from the difference between pre-test and post-test values, which are relativized for the standard error differential (see Table 2).

Table 2

Overview of characteristic values and formulas (3).

| STATISTICAL VALUES AND FORMULAS | VALUES | EXPLANATION |
|---|----------|--|
| σ_x | 0.44 | Standard deviation of test scores PWC ₁₅₀ (pre-test values) |
| ρ_{tt} | 0.94 | Reliability of PWC ₁₅₀ |
| $Z_{1-\alpha}$ | 1.28/90% | Critical z-value of standard normal distribution for a certain error probability α in one-sided testing |
| $RCI=(x_{v2}-x_{v1})/\sigma_{EDiff}$ | | Reliable Change Index x_{v1} -value pre-test x_{v2} -value post-test |
| $\sigma_{EDiff}=\sqrt{(2(\sigma_x)^2)}$ | | Standard error of difference between two test scores |
| $\sigma_E=\sigma_x\sqrt{(1-\rho_{tt})}$ | | Standard error of measurement |
| $VI_u=x_v-\sigma_x\sqrt{(\rho_{tt}(1-\rho_{tt}))} * Z_{1-\alpha}$ | | Confidence interval lower bound (x_v is the estimated true value of the subject) |

Alternative characteristic values for determination of any given intraindividual changes are more complex, modified differential values, such as the characteristic values $V_{\text{descriptive}}$ or V_{infer} (28). These do not, however, bear up against statistical checks with regards to α -error, and as such are likewise not recommended for use in single-case diagnosis of the sports science and training fields (18, 19, 29). In this context, it could be demonstrated that using the standard normal distribution as a test value, the characteristic value V_{infer} is no longer a standard normally-distributed value and undervalues the α -error probability. Ultimately, when using this characteristic value, nothing can be said about the actual size of the error risk in a given individual case (29). When it comes to determining the reliability of any given individual's performance change, e.g. in the training process, preference is to be given to the RCI over any other measured values if the current state of knowledge is to be respected, as the RCI represents the test value for change with the best statistical properties (18, 19, 29).

Furthermore, for assessment of any given intraindividual changes as part of training studies, the variation coefficient is frequently and almost paradigmatically given and recommended (8, 10, 16). However, the variation coefficient reflects the distribution of a characteristic taking group variation at different points in time into consideration, and also proves to be unsuitable for the diagnosis of individual performance progression due to lacking (conventional) limit values and insufficient accuracy (33).

Specific Methodological Considerations

The determination of PWC is used as an established measuring instrument in the field of leisure and health-orientated sport, yielding training and treatment recommendations on the basis of tables of standard values (21, 27). PWC_{150} is the denotation used to describe performance on a bicycle ergometer at a heart rate of 150 beats/min. In order to make a meaningful comparison between all individuals possible, the performance capacity assessment is carried out relative to body mass (9, 15). The load schema compiled (amongst others) by the Federal Committee of Performance Sports of the German Sports Federation [Bundesausschuss Leistungssport des Deutschen Sportbundes] for bicycle ergometry (BAL schema), or the guidelines from the German Cardiology Society [Deutschen Gesellschaft für Kardiologie] (DGK, 2000) (14, 31) serve as the basis for selection of starting load levels, as well as duration of each load and loads for each subsequent level. During the stress tests as carried out, the physical performance capacity of students of sport science who had undergone moderate endurance training were selected for determination of PWC_{150} according to the BAL schema (17).

The concept of the Reliable Change Index (RCI) takes into account (as do all other concepts) the fact that changes cannot be 100% free of measurement errors, and are calculated from the difference between pre-test and post-test values,

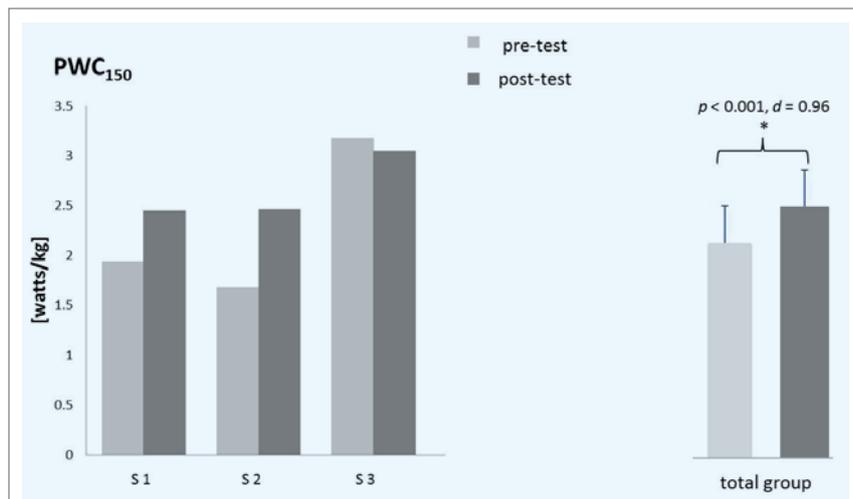


Figure 1

Physical working capacity (PWC_{150}) in pre- and post-test of three individual subjects (S1, S2, S3; left side) and of total group (N=20, Mean \pm SD, right side).

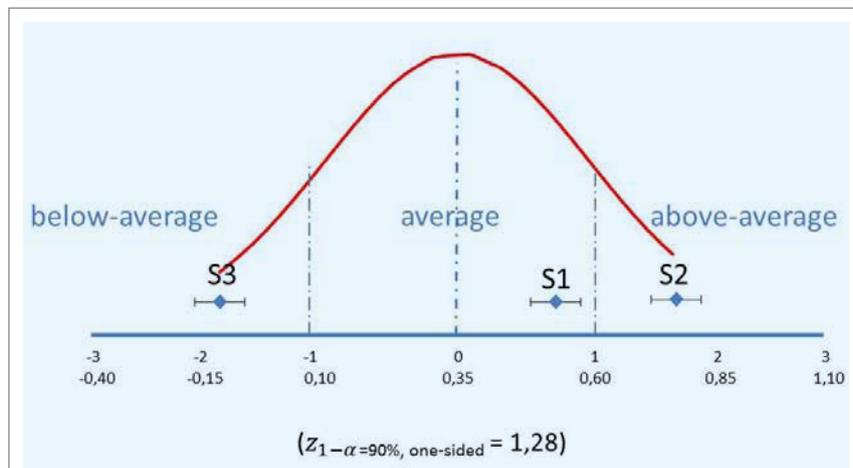


Figure 2

Classification of the individual difference values taking into account the individual confidence intervals in comparison with the group mean value.

which are relativized for the standard error differential (see Table 2). For calculation of standard error, the standard error of measurement is once again employed, taking a reliability value into account, e.g. the reliability from repeating the test, or the split-half reliability method (3). The RCI determined on this basis represents a z-value considering a confidence level or confidence interval determined a priori, which can be compared to the critical z-value from the standard normal distribution. Where the calculated RCI is greater than the critical z-value, an improvement or deterioration can be said to be clinically significant or meaningful in practical terms, with the change not being a result of lacking measurement accuracy (3).

During diagnosis and assessment of individual training effects, one is predominantly concerned with ensuring that no practically significant training effects are being overlooked, and practically insignificant training effects are not being overestimated. In order to minimize this risk, an 80% or 90% confidence level, and with this a greater α -error, should be selected for the determination of the confidence intervals, (and also for the RCI) (6, 11, 21). There are two possibilities available for determination of the confidence interval width (80 or 90%); backing up individual results with the aid of a >

Table 3

Example of calculation for subject 1.

Sample calculation

$$\sigma_{\epsilon} = 0.44\sqrt{(1-0.94)} = 0.10777755$$

$$Q_{EDiff} = \sqrt{(2 * 0.10777755^2)} = 0.15242047$$

$$RCI = (2.44 - 1.94) / 0.15242047 = 3.28039926$$

Table 4

Magnitude of change in RCI ($z_{1-\alpha} = 1.28$; $\alpha = 10\%$) between pre-test and post-test for the three selected test subjects. *Interpretation: A practically significant improvement in performance is not evident.

| | PRE-TEST | POST-TEST | CHANGE INDEX |
|-----------|-------------------------|-------------------------|--------------|
| | REL. PWC ₁₅₀ | REL. PWC ₁₅₀ | RCI |
| S1 | 1.94 | 2.44 | 3.28 |
| S2 | 1.68 | 2.46 | 5.12 |
| S3 | 3.17 | 3.05 | -0.79* |

standard error of measurement or standard estimation error (3). The basis for this is provided by the equivalence and regression hypotheses. The equivalence hypothesis is preferred for application-oriented problems, and assumes that the measurement error remains constant, i.e. is of the same magnitude for each person, and that the feature characteristics vary. On the other hand, the regression hypothesis assumes that the true value for an individual must be estimated from the observed values. Based on the assumption that the observed values for subjects represent a good approximation to the true values, the statistical validation of individual values is carried out using the standard error of measurement (equivalence hypothesis) (3). As it is assumed that an improvement in performance would result over the course of the systematic training program, one-sided hypothesis testing is applied. For the (practical) sports-orientated issue of individual training effects, a confidence level of 90% is established (6, 11, 21).

Further, to generate the confidence intervals, reliability estimates of the measured values are necessary. The reliability estimates are carried out in this example on the basis of a data set of a total of 84 measured values (PWC₁₅₀), which were drawn from a previously-published study on PWC₁₅₀ (16). To determine the reliability, the entire data set was classified into two testing halves according to the odd-even method, by odd and even sequence numbers (26, 30). The subsequent strong parallel correlation is calculated on a pre-to-post-test basis, and the reliability estimated to be $\rho_{tt} = .94$. This estimate of measurement accuracy provides a fundamental basis for evidence of individual effects using the RCI over the course of the training process (3).

Table 5

Magnitude of change in RCI ($z_{1-\alpha} = 1.28$; $\alpha = 10\%$) at different time of measurements in relation to the baseline value. *Interpretation: A practically significant improvement in performance is not evident.

| | PRE-TEST-T24 | | PRE-TEST-T60 | | PRE-TEST-T90 | |
|-----------|-------------------------|-------|-------------------------|------|-------------------------|------|
| | REL. PWC ₁₅₀ | RCI | REL. PWC ₁₅₀ | RCI | REL. PWC ₁₅₀ | RCI |
| S1 | 2.03 | 0.59* | 2.38 | 2.89 | 2.44 | 3.28 |
| S2 | 2.44 | 4.99 | 2.71 | 6.76 | 2.46 | 5.12 |

Table 2 gives an overview of characteristic values and formulas which were used as the basis for RCI calculations and to determine the confidence intervals (which were ascertained using Konfi 2.4, a freeware program for psychometric single-case diagnosis) (3, 25).

For the overall sample (N=20), subjects are sports students of ages ranging from 23-27 years (M=25.3; SD=1.3 with weight M=76.8; SD=6.5 kg). Taking into account the performance requirements placed on students, the PWC₁₅₀ was selected as the submaximal performance test (see introduction).

During a 90-day training study, in which the load level (watts/kg body mass) and duration of load for the test subjects were identical, the bicycle ergometer was used for training during daily training units in accordance with the duration method of maximum 65 and minimum 25 minutes with a peak load of 75% and a minimum of 50% of the maximum performance, with submaximal performance tests (PWC₁₅₀) carried out at defined points in time. Regarding questions relating to specific points in time or to different adaptation reactions on an individual level over the course of the examination, the changes were specifically analyzed at around 3 and 8 weeks. At these points in time, as well as both before and after the examination period, the performance development was verified using sports medicine methodology, and the training load was adjusted for a constant individual stress, using exhaustion tests as the basis (lactate performance diagnostics at the Occupational Medical Center of the company Infraseriv GmbH & Co Hoechst KG Frankfurt).

For the three subjects, the data over the entire examination period were recorded with very few missing data (less than 2%), such that the following targeted expectations should be verified on the basis of the empirical characteristic values:

- Is there a clinically significant or practically meaningful improvement in performance?
- At what discrete point in time over the training process can a clinically significant or practically meaningful improvement in performance be expected?

Results

In the first instance, confidence intervals were determined for the individual performance changes, as were difference values from pre-test to post-test, in which the true value for subjects is to be expected with a given probability. Using this, an initial individual performance estimate is made possible in comparison with the group mean value and the standard deviation. Figure 2 shows how the performance changes of the three selected subjects from figure 1 could be classified within the interindividual comparison of mean performance change of the group. Independent of the fact that subject 3 (S3) demonstrates above-average performance in the post-test, he was not able to improve on the performance achieved in the pre-test. Subject 1 (S1)

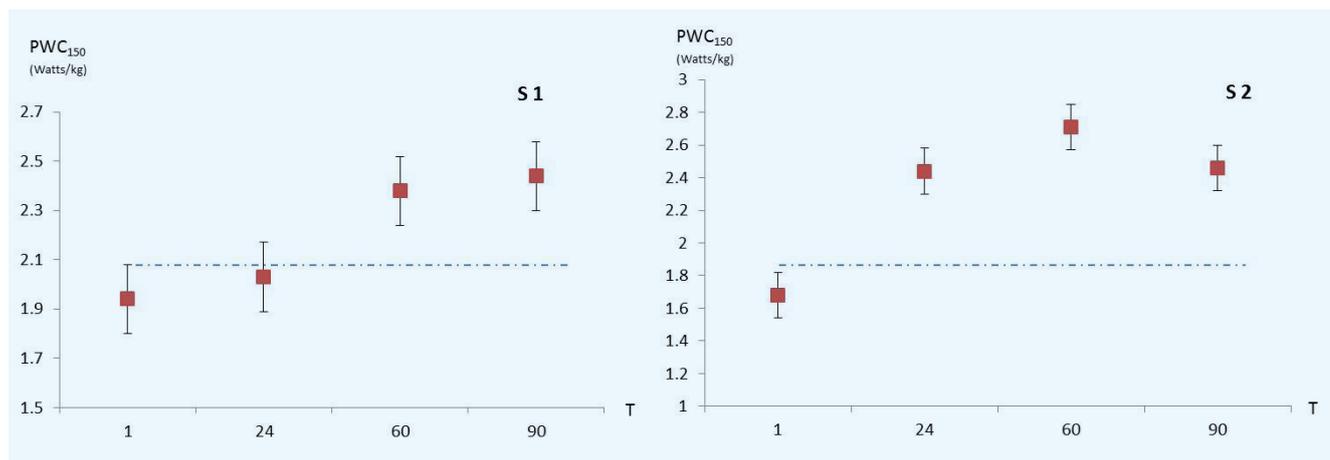


Figure 3

Performance values and confidence intervals at the four time of measurements (T) (observed values) for the examination period for subject 1 (S1, left side) and subject 2 (S2, right side).

was able to significantly improve performance with respect to the pre-test, but despite this remains in the area relating to average performance. The above-average increase in performance of subject 2 (S2) similarly shows average performance in the post-test.

To provide an answer regarding the question of a clinically significant or practically significant training effect from the pre-test to the post-test, in the next step the RCI is calculated. The calculated index (RCI=3.28) for subject 1 is greater than the critical z-value ($z_{1-\alpha}=1.28$; $\alpha=10\%$) and indicates that the changes from pre-test to post-test are not to be traced back to lacking accuracy of measurement, and as such can be interpreted as practically meaningful improvements in performance (see calculation example in Table 3).

In the post-test, subject 2 achieves a value of 2.46 Watts/kg. In this instance too, the calculated RCI=5.12 is greater than the z-value ($z_{1-\alpha}=1.28$; $\alpha=10\%$), indicating that the change from pre-test to post-test is to be interpreted as an individual improvement in performance and thus as a practically meaningful training effect.

For subject 3, the RCI values confirm that no practically significant improvement in performance was achieved over the examination period (see Table 4), such that the other questions for examination shall only be followed up for subjects 1 and 2.

The following table provides an overview of the individual RCI change scores at the points in time when measuring was carried out (henceforth, ‘time of measurement’ [T]) as indicated in the training process, and shows the practically significant improvements in performance at the different measuring time points in relation to the achieved value in the pre-test for subjects 1 and 2.

When considering the changes with respect to each respective previous time of measurements [T] (see Table 6), i.e. from

pre-test to T24, from T24 to T60, and from T60 to T90, an additional practically significant improvement can also be seen between the second (T24) and third (T60) time points for subject 2, in addition to the first practically significant improvement seen from pre-test to measurement time point after 24 days.

In figure 3, the PWC₁₅₀ values at the individual time of measurements (T) as well as the confidence intervals for the individual Ts over the examination period are represented for both of the subjects exhibiting practically significant performance improvements. The dotted guidelines allow clear identification of the different adaptation characteristics of the two subjects.

In summary, it may be stated that subject 1 reacts with a practically significant improvement after just eight weeks (Figure 3, left side) when participating in endurance training over 90 days (using maximum performance to put results into context), whereas subject 2 demonstrates a practically significant improvement in performance at both just three weeks and at the 8-week time points over the examination period (Figure 3, right side).

Discussion

This paper concerns itself with a permanent research question in the field of sports medicine and training science. For example, no significant group effects can be identified, whilst individual subjects do, however, show significant progression in their performance. Similarly, a change in performance can be demonstrated for a group, but the individual improvements or deteriorations in performance may not be derived from this. When considering individual cases, the basic issue presents itself of whether different individual changes in performance also record practically significant training effects or clinically relevant performance change in individual performance >

Table 6

Magnitude of change in RCI ($z_{1-\alpha}=1.28$; $\alpha=10\%$) at different measuring time points in relation to each respective previous time of measurement. *Interpretation: A practically significant improvement in performance is not evident.

| | PRE-TEST–T24 | | | T24–T60 | | | T60–T90 | | |
|----|-------------------------|-------------------------|-------|-------------------------|-------------------------|------|-------------------------|-------------------------|--------|
| | REL. PWC ₁₅₀ | REL. PWC ₁₅₀ | RCI | REL. PWC ₁₅₀ | REL. PWC ₁₅₀ | RCI | REL. PWC ₁₅₀ | REL. PWC ₁₅₀ | RCI |
| S1 | 1.94 | 2.03 | 0.59* | 2.03 | 2.38 | 1.64 | 2.38 | 2.44 | 0.39* |
| S2 | 1.68 | 2.44 | 4.99 | 2.44 | 2.71 | 1.77 | 2.71 | 2.46 | -1.64* |

capacity. The procedures associated with specific and/or restrictive considerations for quantification of individual improvements or deteriorations e.g. time series analysis, are often not employed as methodological usage requirements on these data are not met, and the expectation of more time-consuming and complex analyses can deter from their use. Furthermore, changes can be masked or skewed by effects such as measurement errors and biological variability.

In order to be able to assess the quality of training, the quantity and quality of the changes in each individual case must therefore be recorded reliably and accurately. The Reliable Change Index (RCI) distinguishes between fluctuations due to random error and measurement errors and systematic changes, and furthermore describes the magnitude of changes between pre-test and post-test with the result making a statement on whether changes are statistically and practically significant or not (3). Therein lies a substantial benefit of the RCI due to its reliable determination of clinically significant and/or practically significant performance changes of each individual subject, which in connection with the issue of training effects, are either underestimated or overestimated with a certain confidence level.

For two subjects, practically significant training effects could be demonstrated, which furthermore were distinct, i.e. individualized statements for the point in time of a significant training effect taking place could be formulated. The different time lapses allow individual adaptation characteristics to be derived, and demonstrate the potential of an individual, but also of a monitoring procedure for training processes that is still being (and has until now been) little used in the training process. Practically significant changes that in the case of subject 1 could be seen as early as in week 3, demonstrate that changes to training priorities and/or adaptations to training content could clearly be carried out at an earlier time. On the other hand, an adaptation to training load for subject 2 is indicated only after 8 weeks.

In summary, the examples presented bear witness to the advantages of single-case diagnosis within the training process, as well as to the essential problems encountered by traditional random-sampling-based empirical procedures in connection with identification of individual performance change. The presented single-case diagnosis compensates not only for the methodologically limited significance of group design for leisure time and popular sports, allowing only limited conclusions to be drawn on a given individual subject's adaptation processes, but rather it also uses examples to demonstrate a simple and reliable process that can be employed for individualized training. ■

Conflict of Interest

The authors have no conflict of interest.

References

- (1) **BATTERHAM AM, HOPKINS WG.** Making meaningful inferences about magnitudes. *Int J Sports Physiol Perform.* 2006; 1: 50-57. doi:10.1123/ijsp.1.1.50
- (2) **BOUCHARD C, RANKINEN T.** Individual differences in response to regular physical activity. *Med Sci Sports Exerc.* 2001; 33: S446-S451. doi:10.1097/00005768-200106001-00013
- (3) **BÜHNER M.** Einführung in die Test- und Fragebogenkonstruktion. 3., aktualisierte und erw. Aufl. München: Pearson Studium; 2011.
- (4) **BÜSCH D, STRAUSS B.** Wider die „Sternchenkunde“! *Sportwiss.* 2016; 46: 53-59. doi:10.1007/s12662-015-0376-x.
- (5) **CHELUNE GJ.** Assessing reliable neuropsychological change. In: Franklin R. D., ed. *Prediction in forensic and neuropsychology: Sound statistical practices.* Mahwah, NJ, USA: Lawrence Erlbaum; 2003: 123-147.
- (6) **HAGER W.** Jenseits von Experiment und Quasi-Experiment: Zur Struktur psychologischer Versuche und zur Ableitung von Vorhersagen. Göttingen: Hogrefe; 1992.
- (7) **HAUTALA AJ, KIVINIEMI AM, MAKIKALLIO TH, KINNUNEN H, NISSILÄ S, HUIKURI HV, TULPPO MP.** Individual differences in the responses to endurance and resistance training. *Eur J Appl Physiol.* 2006; 96: 535-542. doi:10.1007/s00421-005-0116-2
- (8) **HECKSTEDEN A, KRAUSHAAR J, SCHARHAG-ROSENBERGER F, THEISEN D, SENN S, MEYER T.** Individual response to exercise training - a statistical perspective. *J Appl Physiol.* 2015; 118: 1450-1459. doi:10.1152/jappphysiol.00714.2014
- (9) **HOLLMANN W, STRÜDER HK, PREDEL H, TAGARAKIS CVM.** Spiroergometrie. Kardiopulmonale Leistungsdiagnostik des Gesunden und Kranken. 1. Aufl. s.l.: Schattauer GmbH Verlag für Medizin und Naturwissenschaften; 2006.
- (10) **HOPKINS WG.** Individual responses made easy. *J Appl Physiol.* 2015; 118: 1444-1446. doi:10.1152/jappphysiol.00098.2015
- (11) **HUBER HP, ED.** Psychometrische Einzelfalldiagnostik. Weinheim: Beltz; 1973.
- (12) **JACOBSON NS, ROBERTS LJ, BERNS SB, MCGLINCHEY JB.** Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J Consult Clin Psychol.* 1999; 67: 300-307. doi:10.1037/0022-006X.67.3.300
- (13) **JACOBSON NS, TRUAX P.** Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991; 59: 12-19. doi:10.1037/0022-006X.59.1.12
- (14) **LÖLLGEN H, HANSEL J, BOLDT F, ET AL; ARBEITSGRUPPE DER DGSP.** S1-Leitlinie Vorsorgeuntersuchung im Sport, 2007, DGSP. www.bayerischer-sportaerzteverband.de/fileadmin/user_upload/html/Download/S1_Leitlinie_DGSP.pdf [7th December 2017].
- (15) **MARÉES HD.** Sportphysiologie. Köln: Sportverlag Strauß; 2003.
- (16) **MARSCHALL F, FERGER K, MÜLLER H.** Biological Variability in Submaximal Parameters of Performance and Strain. *JEPonline.* 2014; 17: 102-112.
- (17) **MEYER T.** Belastungsuntersuchungen: Praktische Durchführung und Interpretation. In: Kindermann W, Dickhuth H, Nief A, Röcker K, Urhausen A, ed. *Sportkardiologie: Körperliche Aktivität bei Herzerkrankungen.* Dordrecht: Springer; 2007: 39-66.
- (18) **NACHTIGALL C, SUHL U.** Warum kompliziert, wenn es auch einfach geht? Teil 1: Zur Analyse intraindividuelle Veränderung. *metheval report.* 2002; 4.
- (19) **NACHTIGALL C, SUHL U.** Evaluation intraindividuelle Veränderung. *Z Klin Psychol Psychother.* 2005; 34: 241-247. doi:10.1026/1616-3443.34.4.241
- (20) **OGLES BM, LUNNEN KM, BONESTEEL K.** Clinical significance: history, application, and current practice. *Clin Psychol Rev.* 2001; 21: 421-446. doi:10.1016/S0272-7358(99)00058-6
- (21) **PETERMANN F.** Veränderungsmessung. Stuttgart: Kohlhammer; 1978.
- (22) **ROST R, HOLLMANN W.** Belastungsuntersuchung in der Praxis. Stuttgart: Thieme; 1982.
- (23) **SCHARHAG-ROSENBERGER F.** Ausmaß, Variabilität und Zeitverlauf von Anpassungserscheinungen an ein 50-wöchiges gesundheitssportliches Ausdauertraining, Saarbrücken, Univ., Diss., 2008.
- (24) **SCHARHAG-ROSENBERGER F, MEYER T.** Ausdauertrainingseffekte: Ergometrische Erfassung und Zusammenhänge mit präventiver Trainingswirkung. *Dtsch Z Sportmed.* 2013; 64: 45-51. doi:10.5960/dzsm.2012.055
- (25) **STEIN C, BÜHNER M.** Konfi: Programm zur psychometrischen Einzelfalldiagnostik: Version 02.04.2004. <http://it-cstein.de>. [23rd June 2017].
- (26) **STELZL I.** Testtheoretische Module. In: Tent L, Stelzl I, Hrsg. *Pädagogisch-psychologische Diagnostik. Band 1: Theoretische und methodische Grundlagen.* Göttingen: Hogrefe Verlag; 1993: 39-201.
- (27) **STEMPER T.** Lehrbuch lizenziierter Fitness-Trainer DSSV. 5. Aufl., überarb. und korr. Hamburg: SSV; 2006.
- (28) **STEYER R, HANNÖVER W, TELSER C, KRIEBEL R.** Zur Evaluation intraindividuelle Veränderung. *Z Klin Psychol.* 1997; 26: 291-299.
- (29) **SUHL U, NACHTIGALL C.** Warum kompliziert, wenn es auch einfach geht? Teil 2: Ergebnisse einer Simulationsstudie zum Vergleich von Veränderungskennwerten. *metheval report.* 2002; 4.
- (30) **TENT L, STELZL I, EDS.** Pädagogisch-psychologische Diagnostik. Band 1: Theoretische und methodische Grundlagen. Göttingen: Hogrefe; 1993.
- (31) **TRAPPE H, LÖLLGEN H.** Leitlinien zur Ergometrie. *Z Kardiol.* 2000; 89: 821-837.
- (32) **WALDORF M, WIEDL KH, SCHÖTTKE H.** On the Concordance of Three Reliable Change Indexes: An Analysis Applying the Dynamic Wisconsin Card Sorting Test. *J Cognit Educ Psychol.* 2009; 8: 63-80. doi:10.1891/1945-8959.8.1.63
- (33) **WISE EA.** Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. *J Pers Assess.* 2004; 82: 50-59. doi:10.1207/s15327752jpa8201_10