Journal of Sports Sciences

ISSN: 0264-0414 (Print) 1466-447X (Online) Journal homepage: http://www.tandfonline.com/loi/rjsp20

A methodological approach to short-term tracking of youth physical fitness: the Oporto Growth, Health and Performance Study

Michele Souza, Joey Eisenmann, Raquel Chaves, Daniel Santos, Sara Pereira, Cláudia Forjaz & José Maia

To cite this article: Michele Souza, Joey Eisenmann, Raquel Chaves, Daniel Santos, Sara Pereira, Cláudia Forjaz & José Maia (2016) A methodological approach to short-term tracking of youth physical fitness: the Oporto Growth, Health and Performance Study, Journal of Sports Sciences, 34:19, 1885-1892, DOI: 10.1080/02640414.2016.1142105

To link to this article: https://doi.org/10.1080/02640414.2016.1142105

JOURNAL OF SPORTS SCIENCES, 2016
VOL. 34, NO. 19, 1885–1892
http://dx.doi.org/10.1080/02640414.2016.1142105

Routledge
Taylor & Francis Group

A methodological approach to short-term tracking of youth physical fitness: the Oporto Growth, Health and Performance Study

Michele Souza[a,b], Joey Eisenmann[c], Raquel Chaves[d], Daniel Santos[e], Sara Pereira[e], Cláudia Forjaz[f] and José Maia[e]

[a]Department of Physical Education, Center of Sports, Federal University of Santa Catarina, Florianópolis, Brazil; [b]CAPES Foundation, Ministry of Education of Brazil, Brasília, Brazil; [c]Department of Radiology, Michigan State University, East Lansing, MI, USA; [d]Department of Physical Education, Federal University of Technology of Paraná, Curitiba, Brazil; [e]CIFI2D, Faculty of Sport, University of Porto, Porto, Portugal; [f]Exercise Hemodynamic Laboratory, School of Physical Education and Sport, University of São Paulo, São Paulo, Brazil

ABSTRACT

In this paper, three different statistical approaches were used to investigate short-term tracking of cardiorespiratory and performance-related physical fitness among adolescents. Data were obtained from the Oporto Growth, Health and Performance Study and comprised 1203 adolescents (549 girls) divided into two age cohorts (10–12 and 12–14 years) followed for three consecutive years, with annual assessment. Cardiorespiratory fitness was assessed with 1-mile run/walk test; 50-yard dash, standing long jump, handgrip, and shuttle run test were used to rate performance-related physical fitness. Tracking was expressed in three different ways: auto-correlations, multilevel modelling with crude and adjusted model (for biological maturation, body mass index, and physical activity), and Cohen's Kappa (*κ*) computed in IBM SPSS 20.0, HLM 7.01 and Longitudinal Data Analysis software, respectively. Tracking of physical fitness components was (1) moderate-to-high when described by auto-correlations; (2) low-to-moderate when crude and adjusted models were used; and (3) low according to Cohen's Kappa (*κ*). These results demonstrate that when describing tracking, different methods should be considered since they provide distinct and more comprehensive views about physical fitness stability patterns.

ARTICLE HISTORY
Accepted 8 January 2016

KEYWORDS
Stability; change; fitness; adolescents

Introduction

Physical fitness (PF) is a complex and multi-faceted construct integrating a wide set of bodily functions – morphological, muscular, motor, cardiorespiratory, and metabolic (Bouchard & Shepard, 1994; Malina, 2001; Ortega, Ruiz, Castillo, & Sjostrom, 2008). Although there is no universal definition, PF is generally regarded as the capacity to perform daily physical activities without undue fatigue (Bouchard & Shepard, 1994; Ortega, Ruiz, & Castillo, 2013; Ortega et al., 2008; Safrit, 1990; Smith et al., 2014). Further, most of the available test batteries include components that are interchangeably applied to assess both performance and health-related PF (Safrit, 1990).

There is a consistent body of evidence relating the favourable effects of moderate-to-high levels of PF to health-related outcomes, namely, adiposity, cardio-metabolic risk factors, bone, and psychological traits in childhood and adolescence (Malina, 2001; Moliner-Urdiales et al., 2010; Ortega et al., 2008, 2011, 2013; Smith et al., 2014). The benefits of PF have also been reported among overweight children and adolescents giving rise to the "fat but fit" paradigm (Brouwer, Stolk, Liem, Lemmink, & Corpeleijn, 2013; Ortega et al., 2008, 2013). Besides health benefits, one of the major reasons to target PF in children and adolescents relates to their developmental periods in terms of acquisition of health behaviours, which are believed to persist later in life (Ortega et al., 2008, 2013; Pate, Oria, & Pillsbury, 2012), i.e., they track over time. From a health promotion perspective, information about the dynamics of short- and long-term changes in PF during childhood and adolescence provides reliable knowledge about the stability in PF allowing the early identification of participants with an unfit profile, that may be useful in the implementation of preventive measures (Da Silva, Beunen, Prista, & Maia, 2013; Janz, Dawson, & Mahoney, 2000; Malina, 2001; Ortega et al., 2013).

Tracking studies are most often conducted to describe patterns of growth or change (Kowalski & Schneiderman, 1992). In epidemiology, tracking is mainly used to estimate the relative stability of factors related to chronic diseases (Kowalski & Schneiderman, 1992, Twisk, 2003b) and to predict, to some extent, later values (Kowalski & Schneiderman, 1992; Malina, 2001, Twisk, 2003b). As such, tracking broadly describes a regular behaviour among a collection of growth patterns, or the tendency of individuals, or collection of individuals, to stay on a particular course of changes in growth, and in our case would be changes in PF (Kowalski & Schneiderman, 1992). Tracking studies on PF generally suggest moderate-to-moderately high coefficients from childhood to adolescence (Da Silva et al., 2013; Malina, 2001; Pahkala et al., 2013) but inter-age auto-correlations can range from low-to-moderately high (e.g., 0.24–0.71 for muscular strength and 0.30–0.61 for cardiorespiratory fitness) (Malina, 2001).

CONTACT Michele Souza ✉ mcsouza85@hotmail.com Department of Physical Education, Center of Sports, Federal University of Santa Catarina, Campus Universitário, 88040-900 Florianópolis, SC, Brazil

Methodologically, there are wealth of statistical techniques to address tracking (Kowalski & Schneiderman, 1992; Twisk, 2003b), but most PF studies are based on auto-correlations (Da Silva et al., 2013; Malina, 2001). Tracking is an important facet when describing longitudinal changes occurring in a certain variable, and in order to take full advantage of the available information, it is of great interest to jointly consider the influence of time-varying and/or time-invariant covariates (Da Silva et al., 2013; Maia et al., 2010; Rodrigues, Leitao, & Lopes, 2013; Twisk, 2003b). The combined use of more adequate statistical techniques makes it possible to examine more challenging issues arising from the notion of (in)stability of longitudinal data that goes beyond the information provided by auto-correlations. For instance, it is possible to explore individual characteristics, dynamics of change of each subject in relation to his/her group over time, as well as inter-individual differences in intra-individual changes. Therefore, the aim of this study is to investigate short-term tracking of cardiorespiratory and performance-related PF among adolescents using three distinct approaches: (1) auto-correlations; (2) multilevel modelling based on a statistical approach suggested by Twisk (2003b) correcting tracking values for time-varying covariates; and (3) Cohen's Kappa (κ) in order to identify group and individual tracking as well as individuals whose trajectories are instable across time. As such, discussing the interpretation of different ways to describe tracking of PF components will enhance our understanding about their (in)stability patterns or changes during the transition from childhood to adolescence. Further, this will provide more reliable information in targeting children at risk when developing more efficient interventions programmes.

Material and methods

Sample

Data are from the Oporto Growth, Health and Performance Study (OGHPS) whose main aim is to investigate, longitudinally, the interaction among individual characteristics, environmental factors and lifestyle predictors that affect growth, development and health aspects of Portuguese adolescents. The OGHPS has a mixed-longitudinal design involving randomly selected adolescents from 10 to 18 years old divided into four cohorts ($n \approx 250–300$ per cohort). The first cohort was followed annually from 10 to 12 years; the second from 12 to 14 years; the third from 14 to 16 years, and the fourth from 16 to 18 years.

For the present study, we will only consider adolescents from the first and second cohorts (10–14 years) given they represent the pubertal period (i.e., mean age at peak height velocity 12 in girls and 14 in boys). To make maximum use of the data, all valid data on each PF test was included. Consequently, sample sizes vary in each PF test (see Table 2). All measurements were conducted annually during the same months. Legal authorisation was obtained from school directors, and parents gave their informed consent; the Ethics Committee of the University of Porto approved the project.

Anthropometry

Anthropometric measurements were made by trained staff following the International Society for the Advancement of Kinanthropometry (Ross & Ward, 1986) standardised protocols. Height was measured to the nearest 0.1 cm with a portable stadiometer (Holtain, UK). Body mass (kg) was measured with a portable bioelectrical impedance scale using the TANITA BC-418 MA Segmental Body Composition Analyser (Tanita, Corporation, Tokyo, Japan) with a 0.1 kg precision. Body mass index (BMI) was computed using the standard formula BMI = [body mass (kg) · (height (m)^2)$^{-1}$].

Biological maturation

Biological maturation was indirectly assessed by the maturity offset regression procedure proposed by Mirwald, Baxter-Jones, Bailey, and Beunen (2002) which estimates how many years a subject is from peak height velocity (PHV). A positive (+) maturity offset represents the number of years the participant is beyond PHV, whereas a negative (–) maturity offset represents the number of years the subject is before from PHV. Estimated age at peak height velocity is determined by chronological age minus maturity offset.

Physical fitness

Aerobic capacity was assessed via 1-mile run/walk test as described in the Fitnessgram battery (Fitnessgram, 1994). In brief, all participants ran/walked a distance of 1609 m in the shortest time possible. Performance-related fitness was assessed with several items from the AAHPERD youth fitness test (AAHPER, 1976) and included: (a) running speed – 50-yard dash test: all participants ran this distance in the shortest time possible; (b) explosive leg power – standing long jump test: all participants jumped as far as possible from a standing position; (c) static strength – grip strength test: all participants gripped the dynamometer (Takei Physical Fitness Test GRIP-D, Japan) with maximum force during 5–10 s; and (d) agility – shuttle-run test: all participants ran as fast as possible from the starting line to a line 9 m away where two small wooden blocks were placed, picked-up one of the blocks, returned to the starting line, placed the block on the line, and then repeated route.

Physical activity

Total physical activity (TPA) was estimated with the Baecke questionnaire (Baecke, Burema, & Frijters, 1982), a reliable and valid instrument (Miller, Freedson, & Kline, 1994; Philippaerts, Westerterp, & Lefevre, 1999) that describes three basic domains of PA, namely, school PA, leisure time PA, and sport participation PA. It comprises a total of 16 questions divided into these domains and each is scored from 1 (minimal PA) to 5 (maximal PA). TPA score is obtained from the unweighted sum of the three domains, and scores ranged from 3 (lowest) to 15 (highest). All adolescents answered the questionnaires during their physical education classes under the supervision of the physical education teachers who were trained by the research team.

Data quality control

Data quality control was assessed in three different steps: (1) training of all team members by experienced researchers of the Kinanthropometry Laboratory of the Sports Faculty, University of Porto, Portugal; (2) conducting random retests (intra-rater reliability) on each assessment day; (3) reliability calculations using the ANOVA-based intra-class correlation coefficient (*R*), as well as the technical error of the measurement (TEM): TEM = 0.1 cm for height and 0.1 kg for body mass; for PF tests, *R* values ranged from 0.88 (shuttle run) to 0.95 (handgrip); for PA, $R = 0.75$.

Statistical analysis

All analyses were stratified by gender. Exploratory and descriptive data analyses were conducted in IBM SPSS 20.0 software to check for the presence of outliers.

Tracking was expressed in three different ways. First, auto-correlations (*r*) were computed in IBM SPSS 20.0 software. Since we have three data points in each cohort, three auto-correlations are available within each cohort. In order to have a single auto-correlation within cohort to express tracking, each value was transformed to a Fisher's *z*-value; then, the three Fisher's *z*-values were added and the mean calculated; finally, the mean Fisher's *z*-value was transformed back into an auto-correlation value, labelled as mean *r* as previously advocated (Beunen et al., 1999; Da Silva et al., 2013).

Second, crude and adjusted tracking coefficients were calculated in HLM 7.01 software, using maximum likelihood parameter estimation techniques, following a statistical model suggested by Twisk (2003b) and expressed by the following formula:

$$Y_{it} = \beta_0 + \beta_1 Y_{it1} + \beta_2 t + \sum_{j=1}^{J} \beta_{3j} X_{ijt} + \varepsilon_{it}.$$

For the sake of clarity, we will define its components by use of an example with the handgrip test. Y_{it} is the observed handgrip level for subject i ($i = 1, \ldots, n$) at time t ($t = 1, \ldots, k$), Y_{it1} is the observed handgrip level at baseline for subject i, β_1 is the regression coefficient used as a tracking coefficient, β_2 is the regression coefficient for time (in our case a 3-year period), X_{ijt} is the time-varying covariate j of individual i (our time-varying variables are BMI, maturity offset, and PA), β_{3j} is the regression coefficient for each time-varying covariate j, J is the number of time-varying covariates, and ε_{it} is the error for subject i at time-point t. The relationships between the initial value at t_1 and the values from t_2 to t_3 were simultaneously estimated, resulting in one single regression coefficient (β_1). Based on this model, crude tracking coefficients were first calculated and the value of the initial measurement at t_1 (Y_{it1}) was regressed on the entire longitudinal data available from t_2 to t_3 (no time-varying predictors were included at this crude model). Then, we calculated adjusted tracking coefficients where Y_{it1} value was also used as in crude model, and additionally the model was corrected for time-varying (X_{ijt}) covariates (BMI, maturity offset and TPA). In order to account for scale differences, all predictor variables were transformed into subgroup specific *z*-scores before entering the model, i.e., subgroup mean centring. The use of standardised variables in the longitudinal analyses makes the coefficients vary between −1 and +1, but since we assume that correlations among the repeated observations are positive, these tracking coefficients have values between 0 and 1, making the tracking coefficients interpretable as a longitudinal correlation coefficient (Twisk, 2003b).

Third, PF tracking was also described following Cohen's Kappa (κ) implemented in Longitudinal Data Analysis (LDA) software (Schneiderman & Kowalski, 1993). This approach is based on the notion that tracking exists if an individual's successive measurements remain in the same quantile of the distribution over time (Kowalski & Schneiderman, 1992). This is a descriptive and nonparametric statistic and following Landis and Koch (1977) values of $\kappa > 0.75$ represent excellent tracking, $0.40 \le \kappa < 0.75$ moderate stability, and $\kappa < 0.40$ poor tracking. Tracking expresses the number of times each subject is within his/her canal, i.e., the greater the number of times, the higher the κ. In this study, we will consider that three distinct longitudinal canals (tertile division of the distribution at each time point) are possible trajectories for all participants given the short duration of the repeated observations. Given that there are three time points and three canals, $3^3 = 27$ individual possible trajectory combinations are possible. From the canals, we identified the percentage of participants whose trajectories were highly instable in the extremes (see Figure 1). Positive instability means participants that at the first

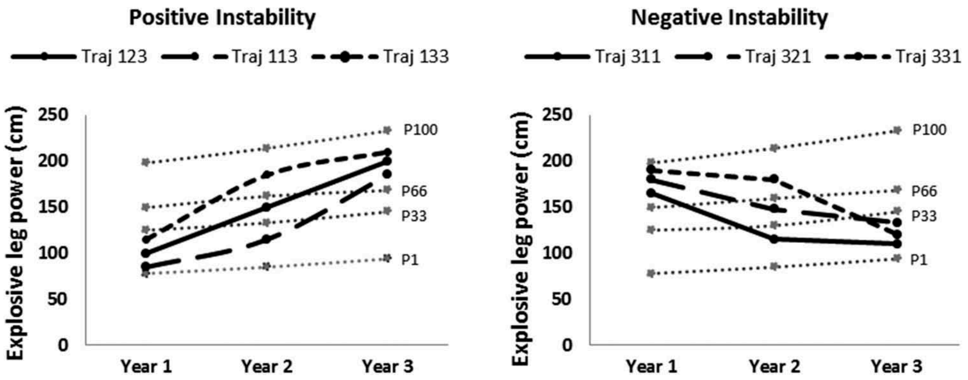

Figure 1. Examples of all possible trajectories of positive and negative instability as previously defined [dotted lines represent percentiles (P1, P33, P66, P100)].

measurement year were in the first canal and then improved their fitness levels reaching the third canal; these are identified as 123, 113, and 133. Negative instability refers to participants that at the first measurement were in the third canal and with time decreased their fitness levels, cross canals and came down to the first canal and are identified as 311, 321, and 331.

Results

Descriptive statistics are shown in Table 1. In general, girls and boys show mean improvements in their PF tests over time. The mean BMI is similar in both sexes. Girls' maturity offset in cohort 1 was closer to PHV and in cohort 2 values were more similar, except at age 12. Girls' and boys' mean TPA values increased over time, but in cohort 2 mean values were more stable.

Table 2 shows the results for tracking. Following cut-points from Malina (2001), auto-correlations show moderate-to-high tracking across cohorts ($0.39 < r_m < 0.79$). Static strength had the highest stability in both cohorts and sexes ($r_m = 0.70$–0.79). On the other hand, cardiorespiratory fitness had the lowest stability in girls from cohort 2 ($r_m = 0.39$).

In the second approach multilevel modelling was used to calculate crude and adjusted stability coefficients, which are, in fact, longitudinal correlations. Using Malina's cut-points, crude coefficients show moderate-to-high tracking and static strength showed more stability over time in boys and girls from both cohorts ($\beta = 0.61$–0.72). However, when tracking was adjusted for dynamic changes in BMI, maturity offset and TPA, β values were lower. In cohort 1, low tracking is evident for cardiorespiratory fitness in both sexes ($\beta = 0.33$ and 0.36 for girls and boys, respectively), and also in boys' agility ($\beta = 0.30$). Speed, explosive leg power and static strength show slightly higher β values. In cohort 2, cardiorespiratory fitness was less stable in girls ($\beta = 0.31$), and speed in boys ($\beta = 0.34$). Both sexes showed "moderate" tracking in the other components.

Cohen's Kappa showed the lowest PF tracking estimates. Following Landis and Koch (Landis & Koch, 1977), κ was poor for all PF components across the cohorts, except for static strength in both cohorts ($0.41 < \kappa < 0.51$). Further, the percentage of positive instability, as previously defined, ranged from 1.4% (girls' velocity in cohort 2) to 7.2% (girls' cardiorespiratory fitness in cohort 1); negative instability ranged from 0.9% (boys' static strength in cohort 2) to 6.0% (girls' cardiorespiratory fitness and boys' agility both from cohort 1).

Discussion

This study used three different approaches to examine the short-term tracking in adolescents' cardiorespiratory and performance-related PF. If a regular pattern is kept over time, then prediction is possible (Kowalski & Schneiderman, 1992; Malina, 2001, Twisk, 2003b). As such, tracking encapsulates two main ideas: stability and predictability (Kowalski & Schneiderman, 1992). Stability is a somewhat elusive concept, such that available statistical techniques to quantify tracking vary and depend on the assumptions linked to the stability-instability concept which may lead to different conclusions (Kowalski & Schneiderman, 1992). For example, in boys' cardiorespiratory fitness of cohort 1 a moderate auto-correlation ($r_m = 0.62$) was found. However, when using the multilevel approach adjusted for time-varying covariates the stability dramatically decreased ($\beta = 0.36$) and remained poor when determined by Cohen's Kappa ($\kappa = 0.37$).

Most of the available results concerning tracking of PA and PF are based on auto-correlations (Malina, 2001). This is a quite simple statistic to calculate and relatively easy to interpret but has known limitations (Malina, 2001, Twisk, 2003b; Twisk, Kemper, & Mellenbergh, 1994). Our results revealed a moderate-to-high PF tracking in the two cohorts. Previous studies reported autocorrelation values ranging from low to moderately high for health-related and performance-related PF. For example, the Muscatine Study (Janz et al., 2000) followed children and adolescents from 10 to 15 years and results were very similar to our findings with stability ranging from 0.39 to 0.44 in cardiorespiratory fitness and from 0.62 to 0.75 in static strength. On the contrary, the Ellisras Longitudinal Study (Monyeki, Koppes, Monyeki, Kemper, & Twisk, 2007) investigated 1-year tracking in 228 boys and 189 girls aged from 11 to 15 years and reported poor stability for agility ($r < 0.07$), low stability for cardiorespiratory fitness ($0.29 < r < 0.36$), and

Table 1. Descriptive statistics for boys and girls of each age cohort.

		Cohort 1			Cohort 2		
		10 years	11 years	12 years	12 years	13 years	14 years
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
1-mile run/walk (min)	♀	10.98 (1.86)	10.41 (1.78)	10.57 (1.82)	10.75 (1.69)	10.10 (1.58)	10.29 (1.76)
	♂	9.57 (1.94)	8.91 (1.88)	9.04 (1.86)	8.99 (1.76)	8.36 (1.75)	7.95 (1.53)
50-yard dash (s)	♀	9.26 (0.84)	8.85 (1.0)	8.51 (0.85)	8.76 (0.97)	8.59 (0.87)	8.33 (0.80)
	♂	8.82 (0.90)	8.52 (1.12)	8.01 (0.99)	8.19 (0.82)	7.97 (0.91)	7.30 (0.75)
Standing long jump (cm)	♀	126.37 (22.12)	134.67 (22.36)	139.12 (21.71)	135.21 (21.79)	144.54 (23.42)	143.41 (23.67)
	♂	137.01 (22.19)	148.57 (22.36)	158.26 (24.61)	153.93 (23.37)	167.67 (27.80)	182.18 (28.60)
Handgrip strength (kg)	♀	17.57 (3.87)	21.20 (4.38)	23.31 (4.14)	22.36 (4.57)	25.15 (4.53)	26.73 (4.49)
	♂	18.29 (4.04)	20.95 (4.60)	25.13 (6.08)	23.34 (5.52)	27.99 (6.68)	32.74 (7.01)
Agility shuttle run (s)	♀	12.63 (1.34)	11.87 (1.27)	11.72 (1.22)	12.04 (1.20)	11.61 (1.21)	11.68 (1.24)
	♂	12.04 (1.43)	11.23 (1.21)	11.02 (1.24)	11.29 (1.12)	10.70 (1.24)	10.35 (1.01)
BMI (kg m^{-2})	♀	20.08 (3.76)	20.37 (3.90)	20.29 (3.73)	20.56 (3.75)	21.33 (3.71)	21.93 (3.62)
	♂	20.09 (3.73)	20.14 (3.71)	20.15 (3.92)	20.23 (3.65)	20.68 (3.53)	21.29 (3.42)
Maturity offset (years)	♀	−1.54 (0.57)	−0.72 (0.57)	0.05 (0.51)	−0.26 (0.59)	0.53 (0.54)	1.17 (0.49)
	♂	−2.24 (0.58)	−1.40 (0.66)	−0.52 (0.77)	−0.92 (0.74)	0.11 (0.82)	1.07 (0.85)
Total physical activity	♀	7.05 (1.09)	7.56 (1.11)	8.33 (1.18)	7.45 (1.33)	7.63 (1.20)	7.60 (1.21)
	♂	7.75 (1.22)	8.31 (1.19)	9.00 (1.21)	8.56 (1.38)	8.55 (1.32)	8.54 (1.22)

SD = standard deviation; min = minutes; s = seconds; cm = centimetres; kg = kilograms; m = meters.

Table 2. Mean auto-correlations, estimated population tracking coefficients (β), and Cohen's Kappa for each PF test in each cohort separately for boys and girls.

			Autocorrelation	Multilevel modelling				Cohen's Kappa (κ)			
		n	Mean r (r_m)	Crude $\beta^{\S}$	95% CI	Adjusted β	95% CI	κ	95% CI	Instability + (%)	Instability − (%)
Cohort 1											
1-mile run/walk (min)	♀	252	0.53	0.46	0.37; 0.55	0.33	0.24; 0.42	0.29	0.24; 0.34	7.2	6.0
	♂	257	0.62	0.54	0.45; 0.63	0.36	0.27; 0.45	0.37	0.32; 0.42	2.4	2.0
50-yard dash	♀	240	0.64	0.60	0.52; 0.68	0.49	0.40; 0.58	0.40	0.35; 0.46	2.9	2.5
	♂	255	0.57	0.61	0.53; 0.69	0.46	0.38; 0.54	0.41	0.35; 0.46	3.2	2.0
Standing long jump	♀	253	0.66	0.58	0.50; 0.66	0.49	0.41; 0.57	0.35	0.30; 0.40	3.2	2.8
	♂	270	0.68	0.63	0.55; 0.71	0.48	0.40; 0.56	0.37	0.32; 0.42	2.5	3.4
Handgrip strength	♀	291	0.73	0.61	0.55; 0.67	0.48	0.42; 0.54	0.42	0.37; 0.47	1.6	2.0
	♂	295	0.74	0.71	0.64; 0.78	0.48	0.40; 0.56	0.42	0.37; 0.47	5.4	3.0
Agility shuttle run	♀	258	0.53	0.51	0.42; 0.60	0.44	0.35; 0.53	0.33	0.28; 0.38	5.4	5.8
	♂	266	0.48	0.37	0.28; 0.46	0.30	0.21; 0.39	0.30	0.25; 0.35	5.7	6.0
Cohort 2											
1-mile run/walk (min)	♀	258	0.39	0.39	0.30; 0.48	0.31	0.22; 0.40	0.22	0.17; 0.27	4.3	5.0
	♂	359	0.53	0.54	0.47; 0.61	0.43	0.36; 0.50	0.32	0.27; 0.36	2.5	3.6
50-yard dash	♀	216	0.63	0.61	0.51; 0.71	0.49	0.40; 0.58	0.36	0.31; 0.42	1.4	3.2
	♂	321	0.53	0.46	0.39; 0.53	0.34	0.26; 0.42	0.31	0.27; 0.36	3.8	3.8
Standing long jump	♀	249	0.61	0.60	0.52; 0.68	0.54	0.46; 0.62	0.31	0.25; 0.36	2.0	3.6
	♂	348	0.63	0.60	0.53; 0.67	0.51	0.44; 0.58	0.44	0.39; 0.48	2.6	2.5
Handgrip strength	♀	246	0.70	0.69	0.62; 0.76	0.57	0.48; 0.66	0.41	0.36; 0.46	2.4	1.6
	♂	344	0.79	0.72	0.67; 0.78	0.48	0.42; 0.54	0.51	0.46; 0.55	1.5	0.9
Agility shuttle run	♀	228	0.44	0.48	0.39; 0.57	0.43	0.34; 0.52	0.27	0.21; 0.32	3.1	3.9
	♂	327	0.54	0.49	0.41; 0.57	0.43	0.35; 0.51	0.31	0.34; 0.45	3.0	3.9

$^{\S}\beta$ represents longitudinal correlations as described previously in the statistical analysis.

moderate for explosive leg power ($0.53 < r < 0.57$). More recently, 294 Brazilian girls from four age cohorts (8, 10, 12, and 14 years) were followed for three consecutive years (Da Silva et al., 2013), and compared to our tracking coefficients were similar in cardiorespiratory fitness ($0.39 < r < 0.50$) and agility ($0.46 < r < 0.59$), but slightly higher for static strength ($r = 0.87$), and explosive leg power ($0.70 < r < 0.73$). Methodological and biological factors may account for some discrepancy in PF tracking (Malina, 2001). Methodologically, the auto-correlational approach does not provide a single tracking value if more than two time points are available [the number of possible auto-correlations is given by $k(k-1)/2$] such as the case in this study and other truly longitudinal investigations. To resolve this problem, we calculated a mean r, but previous studies only reported correlations between the follow-up and preceding years (Janz et al., 2000; Monyeki et al., 2007) or a weighted mean r (Da Silva et al., 2013). This, of course, may partly explain the disparity in auto-correlations. In addition, the likelihood of obtaining higher stability values is very high with shorter follow-up periods (Malina, 2001). Moreover, tracking is likely to be low with younger participants at baseline (Malina, 2001). However, our results did not provide this evidence. The lower stability in cohort 2, mainly in girls' cardiorespiratory fitness, can be partially explained by the fact that, on average, they have already started the adolescent spurt and are closer to their PHV as compared to those from cohort 1. During this period, variability in timing and tempo of the growth spurt associated to maturation-related changes in body mass are likely to affect performance by increasing inter-individual variation in PF development which potentially influence tracking (Malina, 2001). In summary, the tracking statistic mostly used in sport sciences/physical activity epidemiology research is auto-correlation, despite that it may present several shortcomings. Further, there are viable alternatives, more informative and richer in substantive terms.

The multilevel modelling tracking approach as suggested by Twisk (Twisk, Kemper, & van Mechelen, 2000, Twisk, 2003b) is far more informative than auto-correlations given that (1) all available longitudinal data is used to estimate stability coefficients; (2) it is suitable for designs with unequally spaced time intervals; (3) it enables the study of inter-individual differences in intra-individual changes; and (4) relevant time-varying and/or invariant covariates are incorporated as confounder variables into the tracking analysis. As such, the resulting stability coefficient can be interpreted as a longitudinal (partial) correlation coefficient (Kristensen et al., 2008, Twisk et al., 2000, Twisk, 2003b). Crude coefficients were moderate-to-high; however, when tracking was adjusted for dynamic covariates systematic decreases in all fitness components stabilities were found. It is somewhat difficult to compare our results with previous publications because they are mostly analysed with auto-correlations and never include the effects of dynamic covariates in their reports. As reported, the calculation of a tracking index based only on the longitudinal development of a PF component, without considering confounding variables, can highly influence tracking estimates (Twisk et al., 2000), and may affect predictability of future values. PF is part genetically determined (Bouchard, Malina, & Pérusse, 1997), however it has been shown that several other factors can also influence performance during childhood and adolescence, namely biological maturation (generally, positive association) (Da Silva et al., 2013; Freitas et al., 2002; Malina, 2001; Malina, Bouchard, & Bar-Or, 2004), BMI (negative association in motor tasks linked to jumping, running and lifting, but positive association with strength tasks) (Malina et al., 2004), and PA (positive association, although in lower magnitude and varying from study to study) (Da Silva et al., 2013; Malina, 2001; Malina et al., 2004; Parikh & Stratton, 2011). After controlling for time-varying covariates, boys' cardiorespiratory fitness and static strength had a greater reduction in their stability. This may be due to the fact that variation in performance associated to maturation is more noticeable in boys than girls (Malina et al., 2004). Early maturing boys tend to perform systematically better, especially in motor tasks emphasising strength, power, velocity, and aerobic capacity as they are usually taller, heavier, stronger, and consequently they have higher absolute heart and muscle volumes (Malina et al., 2004). Additionally, they are more motivated to excel in physical exercises/sports as compared to less maturing boys (Malina et al., 2004). Low-to-moderate tracking in the others fitness components suggests that, even during this short-term follow-up period, PF is not as stable as reported by auto-correlations (see Figure 2).

Compared to the other statistical methods to determine tracking, Cohen's Kappa provides a distinct view. More specifically, Cohen's Kappa allows to examine the notion of growth canals (or channels), widely used in auxology to better describe and monitor children and adolescents' growth and development processes (Kuczmarski et al., 2000). The results showed the lowest stability estimates of the three tracking methodologies used in this study. Static strength was the only component with moderate tracking in both cohorts. These results indicate that changes in almost all PF components does not occur in specific canals over time which reflects a strong instability of individual trajectories in these cohorts, and affects the accuracy of the predictability for future values.

In this sense, it is possible to suggest that adolescents with a fit profile can remain with these high performance levels or decrease their fitness over a short-term time period. From a health perspective, it is very important to follow those children/adolescents who have highly unstable PF across time (Twisk, 2003a). Our approach to extreme PF pattern changes indicated that the percentage of positive instability was higher in girls' cardiorespiratory fitness in cohort 1 (7.2%); and negative instability was higher for girls' cardiorespiratory fitness, as well as boys' agility (6.0%), both from cohort 1. Overall, these result from Cohen's Kappa suggest that: (1) girls from cohort 1 (10–12 years) were more susceptible to positive and negative short-term changes in cardiorespiratory fitness, and this extreme instability could be related to timing and tempo in their maturation process (Malina et al., 2004); and (2) some boys tended to decrease their performance more in agility compared to other fitness components over time, maybe because agility is a complex fitness task that combines speed, balance, power and coordination (Ortega et al., 2008).

Regardless of the statistical approach used to describe tracking, static strength was the most stable PF component across the cohorts for both sexes. This may be due to the linear increase in strength that occurs during adolescent growth and mainly because handgrip is a quite simple test that does not require some level of motor skills like the other fitness components (Malina et al., 2004), namely, agility. On the other hand, aerobic capacity was the fitness component with lower tracking coefficients, mostly in girls. This instability indicates that aerobic component is not firmly rooted yet, which can be related to several factors, such as, age, body size, and maturity status (Malina et al., 2004). Due to its greater instability, aerobic capacity could be more frequently measured than static strength to better understand its intra-individual changes, inter-individual differences and putative dynamic correlates during the adolescent years.

There are potential limitations in the present study. First, it would be appropriate to use direct laboratory measures to assess cardiorespiratory fitness. However, this would be very difficult to implement in large field studies not only for financial and technical reasons, but also for practicality.

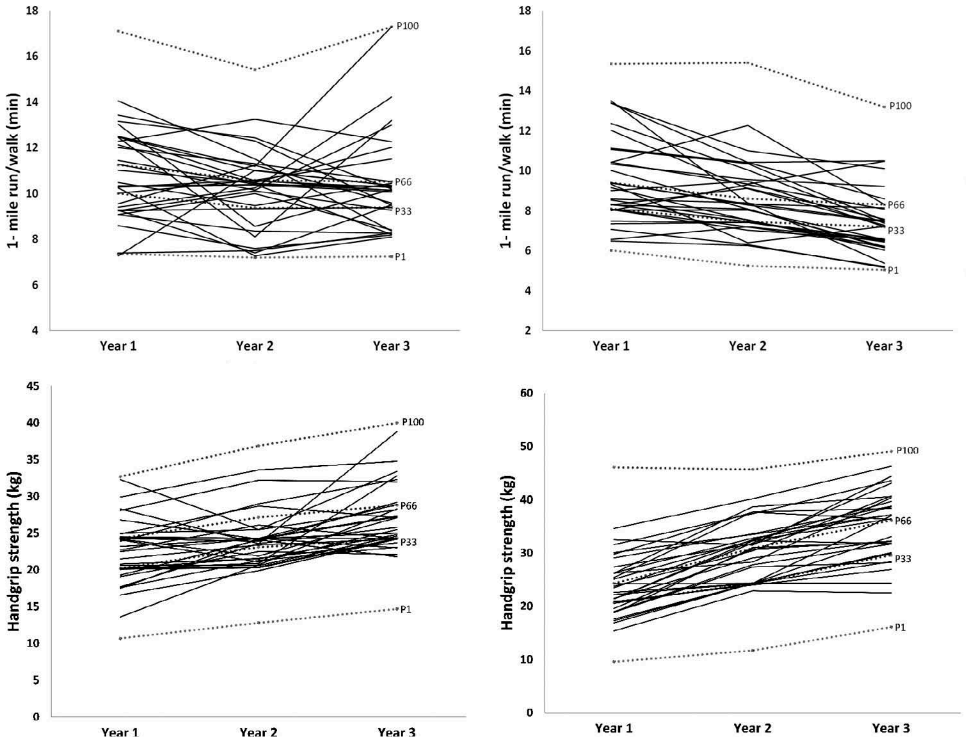

Figure 2. Example of trajectories of 30 random girls (left) and boys (right) from cohort 2 in the 1-mile run/walk and handgrip strength (the other tests follow the same trend of instability).

Second, PA estimates were based on a questionnaire which is prone to errors. However, all children and adolescents completed all questions in highly standardised conditions; further, high reliability estimates have been reported with this questionnaire in Portuguese children and adolescents from different geographical regions (Ferreira, Marques, & Maia, 2002; Vasconcelos & Maia, 2001). Third, there was some difficulty of comparisons of present data with other studies due to the distinct approach to quantify tracking, as well as the shortage of available longitudinal data about adolescents' physical fitness. Despite these limitations, the study has several strengths. First, short-term PF tracking was described in differential terms using distinct statistical approaches. Second, the use of a novel approach allowed for the effects of time-varying covariates. Third, the identification of individual tracking and unstable trajectories was also presented. Fourth, the PF reliability estimates of this study demonstrate the quality control of data collection. Fifth, the size of the sample and its representativeness in age and sex is warranted. Sixth, the age cohorts represent important transition periods from childhood to adolescence.

This paper provides a unique perspective on the use of tracking. Main results show that PF has a moderate-to-high tracking when using auto-correlations. However, when tracking was described by multilevel modelling and Cohen's Kappa, tracking coefficients were low-to-moderate (see Figure 2). These relatively conflicting results reveal that (1) PF tracking was influenced by a set of correlates, namely biological maturation, BMI, and PA; (2) there was instability of individual trajectories and also inter-individual differences in PF intra-individual changes; (3) when tracking research is conducted it is necessary to first define, as clear as possible, what is meant by stability and relative position, keeping in mind that depending on the statistical approach results may be different because the idea of relative position is distinct. It is thus imperative to first begin with a precise research question, namely, what does the researcher wants to answer? The sequential steps suggested by Collins (2006) are still valid and very important: (a) the use of a well-articulated theoretical model of change using (b) a temporal design that affords a clear and detailed view of the process, with the resulting data analysed by (c) a statistical model that mirrors the theoretical model of change. By integrating these three elements, tracking studies are more conducive to success and richer in their meaning.

In conclusion, this paper reflects some concerns regarding the widespread use of auto-correlations as a tracking statistical summary, which has never been the subject of careful consideration in statistical terms. Further, most tracking papers do not consider the dynamics of other co-variables that affect the behaviour of the dependent variable of interest. In sum, it was shown that if the researcher is interested in obtaining simple descriptive statistics, the simplest option is to use auto-correlations. Conversely, if she/he aims to investigate inter-individual differences in intra-individual changes together with the possible effects of relevant time-varying and/or fixed covariates (considered as confounders), then multilevel modelling is probably one of the best approaches. If the interest centres around stability based on the idea of growth canals in order to identify group and individual tracking, the best option may be Cohen's Kappa.

Acknowledgements

We would like to thank the study participants along with their parents, teachers, and school principals for their involvement in the study.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

American Alliance for Health, Physical Education, Recreation and Dance (AAHPER). (1976). *Youth fitness test manual*. Washington, DC: AAHPERD.

Baecke, J., Burema, J., & Frijters, J. (1982). A short questionnaire for the measurenment of habitual physical activity in epidemiological studies. *American Journal of Clinical Nutrition*, *36*(5), 936–942.

Beunen, G., Lefevre, J., Claessens, A., Maes, H., Renson, R., Simons, J., … Sneyers, C. (1999). Tracking in health and performance-related fitness from adolescence to adulthood. In B. Bodzár & C. Susanne (Eds.), *Studies in human biology* (pp. 257–262). Budapest: Ëötvös University Press.

Bouchard, C., Malina, R., & Pérusse, L. (1997). *Genetics of fitness and physical performance*. Champaign, IL: Human Kinetics.

Bouchard, C., & Shepard, K. (1994). Physical activity, fitness and health. The model and the key concepts. In C. Bouchard, K. Shepard, & T. Stephens (Eds.), *Physical activity, fitness and health: International proceedings and consensus statement* (pp. 77–78). Champain, IL: Human Kinetics.

Brouwer, S. I., Stolk, R. P., Liem, E. T., Lemmink, K. A., & Corpeleijn, E. (2013). The role of fitness in the association between fatness and cardiometabolic risk from childhood to adolescence. *Pediatric Diabetes*, *14*(1), 57–65. doi:10.1111/j.1399-5448.2012.00893.x

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*, *57*, 505–528. doi:10.1146/annurev.psych.57.102904.190146

Da Silva, S. P., Beunen, G., Prista, A., & Maia, J. (2013). Short-term tracking of performance and health-related physical fitness in girls: The Healthy Growth in Cariri study. *Journal of Sports Sciences*, *31*(1), 104–113. doi:10.1080/02640414.2012.723817

Ferreira, J., Marques, A., & Maia, J. (2002). *Physical fitness, physical activity and health in young population from Viseu- A study in children and youngsters of both gender from 10 to 18 years old.* Viseu: Departamento Cultural – Instituto Superior Politécnico de Viseu.

Fitnessgram. (1994). *The prudencial fitnessgram test administration manual*. Dallas, TX: The Cooper Institute for Aerobics Research.

Freitas, D., Maia, J., Beunen, G., Lefevre, J., Claessens, A., & Marques, A., … Crespo, M. (2002). *Crescimento Somático, Maturação Biológica, Aptidão Física, Atctividade Física e Estatuto Socioeconómico de Crianças e Adolescentes Madeirenses- O Estudo de Crescimento da Madeira*. Funchal: SAEFD Universidade da Madeira.

Janz, K. F., Dawson, J. D., & Mahoney, L. T. (2000). Tracking physical fitness and physical activity from childhood to adolescence: The muscatine study. *Medicine & Science in Sports & Exercise*, *32*(7), 1250–1257. doi:10.1097/00005768-200007000-00011

Kowalski, C. J., & Schneiderman, E. D. (1992). Tracking: Concepts, methods and tools. *International Journal of Anthropology*, *7*(4), 33–50. doi:10.1007/BF02447868

Kristensen, P. L., Moller, N. C., Korsholm, L., Wedderkopp, N., Andersen, L. B., … Froberg, K. (2008). Tracking of objectively measured physical activity from childhood to adolescence: The European youth heart study. *Scandinavian Journal of Medicine & Science in Sports*, *18*(2), 171–178. doi:10.1111/j.1600-0838.2006.00622.x

Kuczmarski, R., Ogden, C., Grummer-Strawn, L., Flegal, K., Guo, S., Wei, R., … Johnson, C. (2000). *CDC growth charts: United States*. Washington, DC: U.S. Government Printing Office.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310

Maia, J. A., Lefevre, J., Claessens, A. L., Thomis, M. A., Peeters, M. W., & Beunen, G. P. (2010). A growth curve to model changes in sport participation in adolescent boys. *Scandinavian Journal of Medicine and Science in Sports*, *20*(4), 679–685. doi:10.1111/j.1600-0838.2009.00990.x

Malina, R., Bouchard, C., & Bar-Or, C. (2004). *Growth, maturation and physical activity*. 2nd ed. Champaign, IL: Human Kinetics.

Malina, R. M. (2001). Physical activity and fitness: Pathways from childhood to adulthood. *American Journal of Human Biology*, *13*(2), 162–172. doi:10.1002/(ISSN)1520-6300

Miller, D. J., Freedson, P. S., & Kline, G. M. (1994). Comparison of activity levels using the Caltrac accelerometer and five questionnaires. *Medicine & Science in Sports & Exercise*, *26*(3), 376–382. doi:10.1249/00005768-199403000-00016

Mirwald, R. L., Baxter-Jones, A. D., Bailey, D. A., & Beunen, G. P. (2002). An assessment of maturity from anthropometric measurements. *Medicine & Science in Sports & Exercise*, *34*(4), 689–694. doi:10.1097/00005768-200204000-00020

Moliner-Urdiales, D., Ruiz, J. R., Ortega, F. B., Jimenez-Pavon, D., Vicente-Rodriguez, G., Rey-Lopez, J. P., ... Moreno, L. (2010). Secular trends in health-related physical fitness in Spanish adolescents: The AVENA and HELENA studies. *Journal of Science and Medicine in Sport*, *13*(6), 584–588. doi:10.1016/j.jsams.2010.03.004

Monyeki, M. A., Koppes, L. L., Monyeki, K. D., Kemper, H. C., & Twisk, J. W. (2007). Longitudinal relationships between nutritional status, body composition, and physical fitness in rural children of South Africa: The Ellisras longitudinal study. *American Journal of Human Biology*, *19*(4), 551–558. doi:10.1002/ajhb.20616

Ortega, F. B., Artero, E. G., Ruiz, J. R., Espana-Romero, V., Jimenez-Pavon, D., Vicente-Rodriguez, G., … Castillo, M. J. (2011). Physical fitness levels among European adolescents: The HELENA study. *British Journal of Sports Medicine*, *45*(1), 20–29. doi:10.1136/bjsm.2009.062679

Ortega, F. B., Ruiz, J. R., & Castillo, M. J. (2013). Physical activity, physical fitness, and overweight in children and adolescents: Evidence from epidemiologic studies. *Endocrinología Y Nutrición*, *60*(8), 458–469. doi:10.1016/j.endonu.2012.10.006

Ortega, F. B., Ruiz, J. R., Castillo, M. J., & Sjostrom, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity (Lond)*, *32*(1), 1–11. doi:10.1038/sj.ijo.0803774

Pahkala, K., Hernelahti, M., Heinonen, O. J., Raittinen, P., Hakanen, M., Lagstrom, H., … Simell, O. (2013). Body mass index, fitness and physical activity from childhood through adolescence. *British Journal of Sports Medicine*, *47*(2), 71–77. doi:10.1136/bjsports-2011-090704

Parikh, T., & Stratton, G. (2011). Influence of intensity of physical activity on adiposity and cardiorespiratory fitness in 5–18 year olds. *Sports Medicine*, *41*(6), 477–488. doi:10.2165/11588750-000000000-00000

Pate, R., Oria, M., & Pillsbury, L. (2012). *Fitness measures and health outcomes in youth*. Washington, DC: The National Academic Press.

Philippaerts, R. M., Westerterp, K. R., & Lefevre, J. (1999). Doubly labeled water validation of three physical activity questionnaires. *International Journal of Sports Medicine*, *20*(5), 284–289. doi:10.1055/s-2007-971132

Rodrigues, L. P., Leitao, R., & Lopes, V. P. (2013). Physical fitness predicts adiposity longitudinal changes over childhood and adolescence. *Journal of Science and Medicine in Sport*, *16*(2), 118–123. doi:10.1016/j.jsams.2012.06.008

Ross, D., & Ward, R. (1986). Scaling anthropometric data for size and proportionality. In T. Reilly, J. Watkins, & J. Borms (Eds.), *Kinanthropometry III. Commonwealth and international conference on sport, physical education, dance, recreation and health*. New York, NY: E&FN Spon.

Safrit, M. J. (1990). *Introduction to measurement in physical education and exercise*. Boston, MA: Times Mirror/Mosby CollegePublishing.

Schneiderman, E., & Kowalski, C. (1993). *LDA. Software system for longitudinal data analysis. Version 3.2*. Dallas, TX: Baylor College of Dentistry.

Smith, J. J., Eather, N., Morgan, P. J., Plotnikoff, R. C., Faigenbaum, A. D., & Lubans, D. R. (2014). The health benefits of muscular fitness for children and adolescents: A systematic review and meta-analysis. *Sports Medicine*, *44*(9), 1209–1223. doi:10.1007/s40279-014-0196-4

Twisk, J. W. (2003a). *Applied longitudinal data analysis for epidemiology*. London: Cambridge University Press.

Twisk, J. W. (2003b). Tracking. In J. W. Twisk (Ed.), *Applied longitudinal data analysis for epidemiology* (pp. 225–240). London: Cambridge University Press.

Twisk, J. W., Kemper, H. C., & Mellenbergh, G. J. (1994). Mathematical and analytical aspects of tracking. *Epidemiologic Review*, *16*(2), 165–183.

Twisk, J. W., Kemper, H. C., & van Mechelen, W. (2000). Tracking of activity and fitness and the relationship with cardiovascular disease risk factors. *Medicine & Science in Sports & Exercise*, *32*(8), 1455–1461. doi:10.1097/00005768-200008000-00014

Vasconcelos, M., & Maia, J. (2001). Is there a decline in physical activity? A cross-sectional study in children and youngsters of both gender from 10 to 19 years old. *Portuguese Journal Sports Science*, *1*, 44–52.