Genetic and environmental determination of tracking in static strength during adolescence

Maarten W. Peeters, Martine A. Thomis, Hermine H. M. Maes, Gaston P. Beunen, Ruth J. F. Loos, Albrecht L. Claesens, and Robert Vlieetinck. Genetic and environmental determination of tracking in static strength during adolescence. J Appl Physiol 99: 1317–1326, 2005. First published June 2, 2005; doi:10.1152/japplphysiol.00021.2005.—The purpose of this study was to determine whether the observed phenotypic stability in static strength during adolescence, as measured by interage correlations in arm pull, is mainly caused by genetic and/or environmental factors. Subjects were from the Leuven Longitudinal Twin Study (n = 105 pairs, equally divided over 5 zygosity groups). Arm-pull data were aligned on age at peak height velocity to attenuate the temporal fluctuations in interage correlations caused by differences in timing of the adolescent growth spurt. Developmental genetic models were fitted using structural equation modeling. After the data were aligned on age at peak height velocity, the annual interage correlations conformed to a quasi-simplex structure over a 4-yr interval. The best-fitting models included additive genetic and unique environmental sources of variation. Additive genetic factors that already explained a significant amount of variation at previous measurement occasions explained 44.3 and 22.5% of the total variation at the last measurement occasion in boys and girls, respectively. Corresponding values for unique environmental sources of variance are 31.2 and 44.5%, respectively. In conclusion, the observed stability of static strength during adolescence is caused by both stable genetic influences and stable unique environmental influences in boys and girls. Additive genetic factors seem to be the most important source of stability in boys, whereas unique environmental factors appear to be more predominant in girls.

Tracking or stability of static strength during adolescence refers to the maintenance of relative rank or position within a group over time (22) and is found to be moderate to moderately high. Intercorrelations in general range between 0.30 and 0.65 over 5- to 7-yr intervals (13, 22). Correlations in static strength from 13 and 18 yr to 30 yr of age among Belgian men were found to be 0.33 and 0.66, respectively (1). High stability in static strength is found during adulthood (13, 18).

In general, the interage correlations tend to decline as the time between observations increases (13, 21), such that the correlation matrix shows a quasi-simplex structure (21). However, a temporarily more irregular pattern in interage correlations during adolescence is sometimes observed, which is probably caused by differences in tempo and timing of the adolescent growth spurt (2, 13). Controlling for the influence of differences in timing of puberty thus may affect interage correlations, leading to a more stable trait showing the quasi-simplex structure that is expected, implying that the further two measurement occasions are separated in time the lower the interage correlations will be. This might be done by aligning ARP performances of each subject on a biological milestone, such as age at peak height velocity (APHV), which also shortens the “growth interval” of the total sample (2).

The heritability estimates of static strength during adolescence derived from family studies range between 0.30 and 0.58 (see Ref. 6 for a review). Recently, heritability estimates for static elbow flexor strength were estimated at 0.40 and 0.30 at 40° and 100° of flexion, respectively, in the 17- to 36-yr-old male sibling pairs from the Leuven Genes for Muscular Strength project (14). In ~15 twin studies, heritabilities for a form of static strength measurement during childhood and adolescence have been reported (6, 7). Common limitations of most studies are small sample sizes, relatively broad age ranges, and no sex-specific heritability estimates. Also, the combination of various methodological approaches and the lack of reported confidence intervals (CIs) complicate comparisons between the studies. The heritability estimates derived from twin studies range from 0.24 to 0.86. The higher estimates are usually found when a limited age range or a longitudinal design is used. Recent univariate cross-sectional analysis of the data from the Leuven Longitudinal Twin Study (LLTS) (7), which is a sample of 105 same-aged twins fol-

Address for reprint requests and other correspondence: M. W. Peeters, Faculty of Kinesiology and Rehabilitation Sciences, K.U. Leuven, Tervuursevest 101, B-3001 Leuven, Belgium (E-mail: Maarten.Peeters@faber.kuleuven.be).

http://www.jap.org 8750-7587/05 $8.00 Copyright © 2005 the American Physiological Society

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
lowered longitudinally from 10 to 18 yr of age, yielded heritability estimates between 0.44 and 0.83 for boys and 0.52 and 0.77 for girls at the various measurement occasions. The remainder of the variation in static strength in that study was explained by unique environmental causes of variation.

Although it is sometimes assumed that stability in a trait provides clues to the relative influence of genetic factors on the trait (10), this is not necessarily the case. The stability in a trait may be caused by stable genetic as well as stable environmental influences or an interplay of both factors. To our knowledge, the first attempt to study the contribution of genetic sources of variation to the stability of static strength was made by Carmelli and Reed (11). They found that, in an elderly (69–80 yr) male population, 35 and 48% of the phenotypic correlations over a 10-yr follow-up were explained by genetic and shared environmental influences, respectively. Data from the 1981 Canada Fitness Survey and the Campbell’s Survey 7-yr follow-up revealed that 32% of the variation in the change scores of handgrip strength over 7 yr was explained by genetic factors (15). The relationship between the stability of static strength during adolescence and the stability of the underlying genetic and environmental causes of variation in the trait, however, has not yet been studied, nor has the possible effect of differences in timing of the adolescent growth spurt on the stability of the trait been considered. The aim of the present study is therefore to explore whether tracking of the ARP performance during the adolescent period is caused by stable genetic and/or stable environmental influences. Research questions in this study are as follows. J) What sources of variation are needed to explain the variation in ARP during adolescence? 2) Do heritability estimates differ between boys and girls after aligning observations on APHV? 3) Is the stability of the trait caused by genetic or environmental factors or both?

**MATERIALS AND METHODS**

Subjects. Subjects are from the LLTS (20). All twins involved in the LLTS belong to the East Flanders Prospective Twin Survey (19). The twins first participated in the study testing program at the LLTS belong to the East Flanders Prospective Twin Survey (19). The remainder of the variation in static strength in that study was explained by unique environmental causes of variation.

Although it is sometimes assumed that stability in a trait provides clues to the relative influence of genetic factors on the trait (10), this is not necessarily the case. The stability in a trait may be caused by stable genetic as well as stable environmental influences or an interplay of both factors. To our knowledge, the first attempt to study the contribution of genetic sources of variation to the stability of static strength was made by Carmelli and Reed (11). They found that, in an elderly (69–80 yr) male population, 35 and 48% of the phenotypic correlations over a 10-yr follow-up were explained by genetic and shared environmental influences, respectively. Data from the 1981 Canada Fitness Survey and the Campbell’s Survey 7-yr follow-up revealed that 32% of the variation in the change scores of handgrip strength over 7 yr was explained by genetic factors (15). The relationship between the stability of static strength during adolescence and the stability of the underlying genetic and environmental causes of variation in the trait, however, has not yet been studied, nor has the possible effect of differences in timing of the adolescent growth spurt on the stability of the trait been considered. The aim of the present study is therefore to explore whether tracking of the ARP performance during the adolescent period is caused by stable genetic and/or stable environmental influences. Research questions in this study are as follows. J) What sources of variation are needed to explain the variation in ARP during adolescence? 2) Do heritability estimates differ between boys and girls after aligning observations on APHV? 3) Is the stability of the trait caused by genetic or environmental factors or both?

**MATERIALS AND METHODS**

Subjects. Subjects are from the LLTS (20). All twins involved in the LLTS belong to the East Flanders Prospective Twin Survey (19). The twins first participated in the study testing program at the LLTS belong to the East Flanders Prospective Twin Survey (19). The remainder of the variation in static strength in that study was explained by unique environmental causes of variation.

Although it is sometimes assumed that stability in a trait provides clues to the relative influence of genetic factors on the trait (10), this is not necessarily the case. The stability in a trait may be caused by stable genetic as well as stable environmental influences or an interplay of both factors. To our knowledge, the first attempt to study the contribution of genetic sources of variation to the stability of static strength was made by Carmelli and Reed (11). They found that, in an elderly (69–80 yr) male population, 35 and 48% of the phenotypic correlations over a 10-yr follow-up were explained by genetic and shared environmental influences, respectively. Data from the 1981 Canada Fitness Survey and the Campbell’s Survey 7-yr follow-up revealed that 32% of the variation in the change scores of handgrip strength over 7 yr was explained by genetic factors (15). The relationship between the stability of static strength during adolescence and the stability of the underlying genetic and environmental causes of variation in the trait, however, has not yet been studied, nor has the possible effect of differences in timing of the adolescent growth spurt on the stability of the trait been considered. The aim of the present study is therefore to explore whether tracking of the ARP performance during the adolescent period is caused by stable genetic and/or stable environmental influences. Research questions in this study are as follows. J) What sources of variation are needed to explain the variation in ARP during adolescence? 2) Do heritability estimates differ between boys and girls after aligning observations on APHV? 3) Is the stability of the trait caused by genetic or environmental factors or both?

**Statistical analysis.** Preece-Baines model I (30) was fitted to the semiannual stature data of each subject to determine APHV for each individual. With the use of this information, ARP performance for each subject at APHV −1 yr, APHV, APHV + 1, APHV + 2, and APHV + 3 yr was calculated by means of linear interpolation based on the two nearest measurement occasions of the yearly physical fitness assessments.

Interage correlations (Pearson) were calculated for boys and girls separately to determine the tracking of the aligned ARP data. Bivariate normality was checked for the aligned data by calculating the Mardia statistics for multivariate skewness and kurtosis.

To determine the relative contribution of genetic and environmental factors to variation in static strength and to simultaneously allow for the tracking and thus covariation between the consecutive observations, longitudinal path models were fitted to the data (8, 9, 25). First, the assumptions for these models were tested, including a test for normal distribution (Shapiro-Wilk test), and a significance test for differences in means (t-test) and variances (F-test) in birth order and zyosity. Mx, a structural equation modeling package (24), was used to compute the goodness of fit of the models and maximum likelihood estimates of the path coefficients. The raw data were used as input in Mx. In this approach, twice the negative log likelihoods of all separate observations (raw maximum likelihood) is obtained and summed, making use of all available data, rather than using the variance-covariance matrices as input, which would result in the loss of information on all pairs with one or more missing data points.

In structural equation modeling, the structural linear equations can be visually represented in path diagrams (e.g., Fig. 1). In these diagrams or models, the latent variables are enclosed in circles. In the “classical” twin study, with data of monozygotic (MZ) and dizygotic (DZ) twins reared together, these latent variables can be additive genetic (A), unique (i.e., nonshared, specific) environmental (E), and common (i.e., shared, familial) environmental (C), or dominant genetic (D) factors. They are unmeasured variables and are the putative causes of variation in and covariation between the observed variables (i.e., ARP at a given age), which are enclosed in squares. The causal paths between the latent variables and observed variables are specified and are depicted as single-headed arrows. The correlational paths, between the latent variables, are depicted by two-headed arrows. Because MZ twins are genetically identical, the correlation between their A factors is 1.0. The correlation between their D factors also equals 1.0. For DZ twins, which like other siblings on average share 50% of their genetic material, the corresponding values are 0.5 for A and 0.25 for D factors. The correlation between the C factors, which cause the members of the pair to be more alike, is 1.0 by definition. There is no correlation between E factors. In the classical twin study, the ACDE model is not identified such that the presence of C and D cannot be tested simultaneously in the same model (8, 25).

To test the hypotheses of the present study, a predetermined strategy testing-specific hypotheses was followed. The models fitted were all simplex models, described by Boomsma and Molenaar (9). In this type of model (e.g., Fig. 1), “innovation” sources of variance (e.g., $A_1-A_3$) explain variation at a given measurement occasion, which is then in part or totally “transmitted” through the transmission paths to the subsequent measurement occasions (but not the previous), explaining a certain amount of variance. This type of model is consistent with data that show a quasi-simplex structure in their interage correlations, implying that the further two measurement occasions are separated in time, the lower their correlations will be. The higher the amount of variance transmitted from previous measurement occasions, the higher the correlations and hence stability of the trait. Residual time-specific (i.e., nontransmitted) variance is modeled as well [time-specific sources of variation, residual additive genetic factor (Ar), residual unique environmental factor (Er), residual common environmental factor (Cr), and/or residual dominant genetic factor (Dr)]. The latter sources of variation are constrained to be equal across time to be distinguishable from the “innovation” sources of...
The set of models tested thus included NSc sex-limitation models, which are the most general type of sex-limitation model, was tested to determine what combination of sources of variance (A, E, C, D) were needed to explain variation in ARP (see Table 4). Subsequently, the sources of variation needed to explain the covariation between the subsequent observations were tested (see Table 4).

The strategy used in model fitting and dropping of the parameters was as follows. First, a set of nonscalar (NSc) sex-limitation models, which are the most general type of sex-limitation model, was tested to determine what combination of sources of variance (A, E, C, D) were needed to explain variation in ARP (see Table 4). Subsequently, nested models of the favored NSc model, including the appropriate sources of variance, were tested to determine, if present, the nature of sex differences [NSc, specific scalar (SS), general scalar, and models without sex-specific parameter estimates]. Subsequently, the sources of variation needed to explain the covariation [e.g., A transmission (At) or E transmission (Et)] between the subsequent observations were tested (see Table 4).

The set of models tested thus included NSc sex-limitation models, which allow nonidentical sets of genes to cause variation in ARP in men and women. In NSc models, the genetic correlation (\(\alpha\)) in all DZ twins is fixed to 0.5, implying that the same set of genes influences the trait in both men and women. The magnitude of the effects of A, E, and C or D, however, does not have to be the same across the sexes. In the SS sex-limitation models, the genetic correlation (\(\alpha\)) in all DZ twins is fixed to 0.5, implying that the same set of genes influences the trait in both men and women. The magnitude of the effects of A, E, and C or D, however, does not have to be the same across the sexes. In the SS sex-limitation models, the genetic correlation (\(\alpha\)) in all DZ twins is fixed to 0.5, implying that the same set of genes influences the trait in both men and women. The magnitude of the effects of A, E, and C or D, however, does not have to be the same across the sexes. Only a general scalar difference is allowed to accommodate a difference in total variance between both sexes. These models result in equal heritabilities for men and women. In the most stringent models, all parameters are constrained to be equal across sexes (25). To determine what sources of variance were needed to explain the covariation between the subsequent observations and hence the stability of ARP performance in this longitudinal study, the transmission paths of the A and E sources of variance were dropped alternatively. The percentages of variance in ARP scores explained by the different latent variables were calculated based on the parameter estimates of the best-fitting and most parsimonious model. The 95% maximum likelihood CIs were estimated (26).

RESULTS

The average raw ARP scores from 10 to 18 yr are given in Fig. 2, A and B, for boys and girls, respectively. In both boys and girls, the ARP performance increases over the whole age period with a clear growth spurt in performance in boys. At the first measurement occasion, ARP data were available for 105 variance. The environmental residuals (Er and Cr) can also account for the measurement errors that inevitably occur when strength is measured. The inclusion of the residual sources of variance allows the phenotypic correlation matrix to deviate from the perfect simplex structure modeled by the innovation and transmission paths and to conform to a quasi-simplex structure.

In the present analyses, the fit of all tested models is compared with that of the saturated model, which provides a baseline fit. The fit of these models is evaluated by the likelihood ratio test (LRT) and a parsimony-based index: Akaike’s information criterion [AIC = (diff. – 2lnL) – (2diff.df)], where diff. – 2lnL is the difference in the −2 times log likelihoods between the genetic and the saturated model, and diff.df is the difference in the degrees of freedom of the two models. As can be seen from the AIC formula, this index favors a more simple model that has fewer estimated parameters and hence more degrees of freedom and a lower value of AIC over a more complex model.

The strategy used in model fitting and dropping of the parameters was as follows. First, a set of nonscalar (NSc) sex-limitation models, which are the most general type of sex-limitation model, was tested to determine what combination of sources of variance (A, E, C, D) were needed to explain variation in ARP (see Table 4). Subsequently, nested models of the favored NSc model, including the appropriate sources of variance, were tested to determine, if present, the nature of sex differences [NSc, specific scalar (SS), general scalar, and models without sex-specific parameter estimates]. Subsequently, the sources of variation needed to explain the covariation [e.g., A transmission (At) or E transmission (Et)] between the subsequent observations were tested (see Table 4).

The set of models tested thus included NSc sex-limitation models, which allow nonidentical sets of genes to cause variation in ARP in men and women. In NSc models, the genetic correlation (\(\alpha\)) in DZ opposite-sex twins is estimated freely between 0.0 and 0.5; the specific (E) and common environmental (C) factors are assumed to be the same for both sexes, but their magnitude may differ. This allows the absolute and relative contribution of A, E, and C or D, to differ between sexes. In the SS sex-limitation models, the genetic correlation (\(\alpha\)) in all DZ twins is fixed to 0.5, implying that the same set of genes influences the trait in both men and women. The magnitude of the effects of A, E, and C or D, however, does not have to be the same across the sexes, again allowing absolute and relative differences in the contribution of the various sources of variance between boys and girls. When A is dropped from the NSc models (e.g., CE model), the model strictly becomes a SS model since \(\alpha\) is no longer included in the model. In the general scalar models, all parameters are set equal across sexes. Only a general scalar difference is allowed to accommodate a difference in total variance between both sexes. These models result in equal heritabilities for men and women. In the most stringent models, all parameters are constrained to be equal across sexes (25). To determine what sources of variance were needed to explain the covariation between the subsequent observations and hence the stability of ARP performance in this longitudinal study, the transmission paths of the A and E sources of variance were dropped alternatively. The percentages of variance in ARP scores explained by the different latent variables were calculated based on the parameter estimates of the best-fitting and most parsimonious model. The 95% maximum likelihood CIs were estimated (26).

RESULTS

The average raw ARP scores from 10 to 18 yr are given in Fig. 2, A and B, for boys and girls, respectively. In both boys and girls, the ARP performance increases over the whole age period with a clear growth spurt in performance in boys. At the first measurement occasion, ARP data were available for 105
boys and 103 girls. At the last measurement occasion, ARP data were available for 91 boys and 87 girls, representing a drop-out rate of 13.3 and 15.5% over the 8 yr of follow-up for boys and girls, respectively. No significant differences in drop-out percentages of the total variance explained by the different models were sufficiently met. The twin correlations at each measurement occasion per zygosity group are provided in Table 3. The cross-twin-cross-measurement occasion correlations, which are not included in this table, in general were higher in MZ twins than in DZ twins as well. The fit of all models tested was significantly worse (LRT: $P < 0.05$) than the baseline fit provided by the saturated model (Table 4), in which no specific structure is imposed on the data, which is probably due to the relatively small sample size and multivariate structure of the data. By means of the LRT, it was determined that dropping C and D from the NSc-ACE and NSc-ADE models, respectively, did not significantly worsen the fit. Dropping A (SS-CE and SS-E models), however, significantly reduced the fit of the model. Based on both the LRT and AIC, the SS-AE model (Fig. 1) was selected as the best-fitting and most parsimonious model. Equating the E or the A sources of variance for boys and girls in the SSA-AE and SSE-AE models also significantly worsened the fit. The transmission could not be restricted to either A or E, as was tested in the SS-AiEi [SS-AE model including innovation (i), transmission (t), and residual (r) additive genetic sources of variance and innovation (i) unique environmental sources of variance] and SS-AiEit models [SS-AE model including innovation (i) additive genetic sources of variance and innovation (i) unique environmental sources of variance]. Both the general scalar-AE and the AE model fitted the data significantly worse than the NSc-AE model as well. The absolute amount of variance explained by the different sources of variation under the SS-AE model is represented in Fig. 3. The 95% CIs on the total A variances obtained by summing At, Ai, and Ar, as well as the percentages of the total variance explained by the different sources of variation under the SS-AE model is presented in Table 5.
sources are represented. Heritability estimates, obtained by summing the percentages of At, Ai, and Ar, are fairly consistent at the various measurement occasions in boys and range between 51.8% (CI 25.7–71.3%) and 81.5% (CI 62.1–90.5%). Genetic factors that already played a role at APHV − 1 still explain 14.5% of the variance at APHV + 3 in boys. The total amount of variance explained by transmission of all previous genetic sources at APHV + 3 adds up to 44.2%. Corresponding values for the Et are 3.5 and 31.2%, respectively. For girls, the range of the heritability is quite large, ranging from 7.4% (CI 0.0–41.3%) to 75.3% (CI 40.0–85.9%). Additive genetic variance transmitted from APHV − 1 explains 15.5% of the variance at APHV + 3. The total amount of variance explained by transmitted A variance at APHV + 3 is 22.4%. Corresponding values for the unique environment are 0.6 and 44.5%, respectively.

DISCUSSION

This is the first study attempting to relate the relative stability of static strength during adolescence at the phenotypic level to the potential stability of the underlying genetic and environmental causes of variation and simultaneously correcting for variation caused by differences in the timing of the adolescent growth spurt.

Since the concept of heritability describes the extent to which differences in a phenotype are explained by genetic differences in a certain population at a certain time (29), generalizations toward the total population can only be made to the extent that the sample under study is representative of this total population. For ARP, it can be seen in Fig. 2B that, in girls, singletons slightly outperform the twins and the differences tend to increase with age. In boys, the twins slightly outperform the singletons between 13 and 15 yr of age, after which the situation is reversed. In both boys and girls, the results of the singletons do fall within one standard deviation of the twin results. It can be concluded that, within the age range used in the present study, the twins correspond relatively well to the reference values of the Belgian population (4, 33). The average APHV of 14.15 yr (SD 0.98) in boys and 12.22 yr (SD 1.11) in girls correspond well to those found in other European and North American samples, which also used Preece-Baines model I to determine APHV (2, 23).

It has been shown that early maturing boys and girls outperform average and late maturing boys and girls on measures of static strength during adolescence (6, 17, 23). In boys, the differences in static strength between early, average, and late maturing subjects have disappeared at age 30 (17). In girls, the associations between skeletal age and static strength observed in early adolescence tend to decline toward late adolescence (3, 23). In the present study, the same relation between maturity status, as assessed by age at peak height velocity, and static strength was observed in both boys and girls (data not shown). The highest correlations were found at the observation closest

Table 2. Interage correlations (tracking) of ARP before and after alignment on APHV

<table>
<thead>
<tr>
<th>Age</th>
<th>10.39</th>
<th>11.42</th>
<th>12.47</th>
<th>13.50</th>
<th>14.53</th>
<th>15.53</th>
<th>16.56</th>
<th>18.41</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.39</td>
<td>1</td>
<td>0.58</td>
<td>0.55</td>
<td>0.45</td>
<td>0.55</td>
<td>0.47</td>
<td>0.43</td>
<td>0.47</td>
</tr>
<tr>
<td>11.42</td>
<td>0.59</td>
<td>1</td>
<td>0.63</td>
<td>0.58</td>
<td>0.59</td>
<td>0.58</td>
<td>0.55</td>
<td>0.38</td>
</tr>
<tr>
<td>12.47</td>
<td>0.29</td>
<td>0.47</td>
<td>1</td>
<td>0.54</td>
<td>0.52</td>
<td>0.57</td>
<td>0.46</td>
<td>0.35</td>
</tr>
<tr>
<td>13.50</td>
<td>0.21</td>
<td>0.29</td>
<td>0.41</td>
<td>1</td>
<td>0.63</td>
<td>0.67</td>
<td>0.52</td>
<td>0.42</td>
</tr>
<tr>
<td>14.53</td>
<td>0.29</td>
<td>0.34</td>
<td>0.51</td>
<td>0.45</td>
<td>1</td>
<td>0.72</td>
<td>0.56</td>
<td>0.53</td>
</tr>
<tr>
<td>15.53</td>
<td>0.05</td>
<td>0.11</td>
<td>0.32</td>
<td>0.53</td>
<td>0.51</td>
<td>1</td>
<td>0.75</td>
<td>0.61</td>
</tr>
<tr>
<td>16.56</td>
<td>0.10</td>
<td>0.16</td>
<td>0.29</td>
<td>0.49</td>
<td>0.42</td>
<td>0.47</td>
<td>1</td>
<td>0.65</td>
</tr>
<tr>
<td>17.49</td>
<td>0.20</td>
<td>0.20</td>
<td>0.29</td>
<td>0.50</td>
<td>0.45</td>
<td>0.43</td>
<td>0.69</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3. Twin correlations for ARP after alignment on APHV

<table>
<thead>
<tr>
<th>Age</th>
<th>MZM</th>
<th>MZF</th>
<th>DZM</th>
<th>DZF</th>
<th>DZO</th>
</tr>
</thead>
<tbody>
<tr>
<td>APHV − 1 yr</td>
<td>0.78</td>
<td>0.60</td>
<td>0.25</td>
<td>0.39</td>
<td>0.27</td>
</tr>
<tr>
<td>APHV</td>
<td>0.85</td>
<td>0.66</td>
<td>−0.29</td>
<td>0.58</td>
<td>0.37</td>
</tr>
<tr>
<td>APHV + 1 yr</td>
<td>0.58</td>
<td>0.21</td>
<td>−0.06</td>
<td>0.12</td>
<td>0.37</td>
</tr>
<tr>
<td>APHV + 2 yr</td>
<td>0.70</td>
<td>0.59</td>
<td>0.13</td>
<td>0.32</td>
<td>0.37</td>
</tr>
<tr>
<td>APHV + 3 yr</td>
<td>0.33</td>
<td>0.70</td>
<td>0.30</td>
<td>0.08</td>
<td>0.01</td>
</tr>
</tbody>
</table>

DZO, dizygotic opposite-sex twins. Note that correlations are only cross-twin-cross-measurement occasion.
to the average APHV, −0.61 for boys and −0.36 for girls, indicating that early maturing adolescents (i.e., having a low APHV) had higher scores on the ARP test. The relationship was stronger in boys than in girls, which is also in agreement with the literature (2). In both boys and girls, the strength of the correlations declined as they got older and only remained borderline significant in boys at age 18.41 yr ($r = −0.24$, $P = 0.02$), which is similar to the results reported by Lefèvre et al. (17). In girls, low negative correlations were found from 13.50 to 15.54 yr of age, after which they approached zero, which is in agreement with Beunen et al. (3) and Malina et al. (23). Because the maturity-strength association may both confound the interage correlations, causing instability in the trait during adolescence, as well as increase the variance around the period of the growth spurt, it was decided to align the ARP performance on APHV. This alignment significantly reduced the variances at the first three measurement occasions in boys and the first measurement occasion in girls ($P < 0.05$) (results not shown). Although body weight shows the strongest association of all body dimensions with strength during childhood and adolescence (5, 16), even after controlling for maturity differences (6), it was decided not to align the data on age at peak weight velocity. This was done because it proved to be impossible for a number of subjects to fit the Preece-Baines curve to the weight data because of fluctuations in body weight in the semiannual observations, especially in girls. Moreover, in the Leuven Growth Study of Belgian Boys (4), the correlation between the age at peak spurt in ARB performance with APHV was slightly higher ($r = 0.38$) than with age at peak weight velocity ($r = 0.32$). It was decided to limit the analysis of the aligned data to the age range of APHV − 1 to APHV + 3 because sample sizes were drastically reduced beyond that range due to lack of measurements for the somewhat earlier maturing girls and somewhat later maturing boys for APHV − 2 and APHV + 4, respectively. In girls, e.g., APHV − 2 would on average equal 10.22 yr of chronological age, whereas the average age at the first measurement occasion of the girls was 10.41 yr (Table 2). Phenotypic tracking of static strength after aligning the data on APHV showed declining correlations with an increasing age interval (further away from the diagonal) (Table 2), thus conforming rather well to a quasi-simplex structure. This was not the case for the raw data, which is in agreement with the results reported by Fortier et al. (13). Maia et al. (21), however, found no such “disturbance” in the interage correlations for static strength during adolescence in boys to which a quasi-simplex structure was fitted. This apparent discrepancy can probably be explained by the age range studied by Maia et al. (21) (12.8–17.7 yr of age), excluding the very early maturing boys, which may have had an effect on the interage correlations. As could be expected, the alignment did not markedly alter the tracking between the first and last measurement occasions, although the interval in the aligned data is shorter (4 yr) than in the unaligned data (±8 yr). This is because aligning the data shortens the growth period (2) of the total sample since all subjects are measured on the same point in time along the way to adult stature. In the unaligned data, however, an early-maturing subject may be 3 yr past his/her APHV, whereas another late-maturing subject may still not have reached his/her APHV, which thus causes the growth period of the overall sample to be longer and leads to more fluctuating interage correlations between the measurement occasions surrounding the adolescent growth spurt. For boys, the tracking over the entire age range in both aligned and unaligned data was within the range of 0.30 to 0.65 for static strength reported in the literature (13, 21, 22). In girls, the tracking in the present study was somewhat beneath that range when the 4-yr interval was considered in the aligned data and markedly lower in the unaligned data.

The fit of all models tested was significantly worse than the baseline fit provided by the saturated model in which no specific structure is imposed on the data, which is probably due to the relatively small sample size and multivariate structure of...
the data. This does warrant some caution for the interpretation of the results. A strength of the present report, however, is that maximum likelihood CIs on the heritability estimates are reported. Other adolescent twin studies on motor performance, which all use comparable or even smaller sample sizes, in general do not report CIs on the heritability estimates. Because phenotypically there is a quasi-simplex structure present in the data, the alternative of fitting independent pathway (IP) models, which include common factors shared by all measurement occasions and thus allowing for covariation via a pleiotropic effect, is probably not appropriate. This type of model in fact parallels a latent factor structure as in a confirmatory factor analysis, which is suggested to be fundamentally unsuited to data conforming to a (quasi-)simplex structure (9). After the initial analysis (Table 4), a SS-AE IP model was fitted to the data to verify whether indeed the fit of this type of model would be worse than the fit of the simplex models. As suggested by Eaves and colleagues (12), the fit of the simplex model was compared with that of the IP model by alternatingly dropping the transmission paths and the IP paths from a model, including both the IP structure and the simplex structure (full model; Fig. 4). These analyses revealed that the IP model fitted the data significantly worse than the full model (diff. \( -2\ln L = 47, 39; \text{diff.df} = 20; P < 0.05 \)). In addition to dropping all transmission paths (paths 6–9, 16–19, 26–29, and 36–39)(Fig. 4), the residual paths (paths 10, 20, 30, and 40) were dropped as well, since, without the transmission paths included, the innovation paths can account for all time-specific variance that is not shared between measurement occasions. Dropping the IP paths from the full model (paths 41–60; Fig. 4) did not significantly worsen the fit (diff. \( -2\ln L = 27, 40; \text{diff.df} = 20; P = 0.12 \)), thus confirming that the simplex model is the most appropriate model to retrieve the phenotypic quasi-simplex structure observed in the present data. The fit was also significantly worsened by simultaneously dropping the IP structure for A (paths 41–45 and 51–55) and the simplex structure for E (paths 16–20 and 36–40) and vice versa, indicating that for both A and E the simplex structure was
needed to adequately retrieve the phenotypic quasi-simplex structure of the data.

The SS-AE simplex model (Fig. 1) was the best-fitting (LRT) and most parsimonious (AIC) model (Table 4). This means that the same A and E factors cause the variation and stability in static strength during adolescence in boys and girls, but the extent to which these factors contribute to the variation in static strength differs significantly between sexes. No evidence for a significant amount of shared environmental or genetic dominance factors was found, which is in agreement with the previous cross-sectional analysis on the nonaligned data (7). Although multivariate analyses of correlated variables increases the statistical power to detect C and D (32), some caution is still warranted here since simulations revealed statistical power in the present analysis of 65% to detect C factors and hence resulting in lower heritability estimates.

In Table 5, it is shown that genetic factors that explained an important amount of variance at the first measurement occasion (Ai) still explain ~15% of the variance at APHV + 3 in both boys and girls. Corresponding values for unique environment are 3.5 and 0.6% in boys and girls, respectively. This indicates that, when the entire age range is considered, genetic factors account for the majority of the stable variance across the adolescent growth period. This genetic stability may be functionally related to the tracking of body mass (23) and, more specifically, fat-free mass. In adult men, fat-free mass has been shown to be highly heritable (34) and the most important determinant of static strength (14), although some caution may be warranted extrapolating these findings to the adolescent period. In boys, a substantial amount of new A variance is introduced at each subsequent measurement occasion (Ai) and then transmitted to the next observations, resulting in a total amount of At of 44.2% at the last measurement occasion. In girls, however, virtually no new A variation is introduced from APHV to APHV + 2 yr are quite low and even not significant at APHV + 2 yr and APHV + 2 yr are quite low and even not significant at APHV + 2 yr, since the lower CI includes 0. The upper limits of the maximum-likelihood CIs do overlap with theheritabilities reported in the literature. Nevertheless the values are also considerably lower than those found for the univariate cross-sectional analysis on the unaligned data of the LLTS reported earlier (7). One might speculate that a partial explanation can possibly be found in the relation between body weight and static strength on one side (6) and the fluctuation in the semianual body weight measurements observed in many girls after the adolescent growth spurt in our sample on the other side. This fluctuation in body weight is not likely to be due to environmental factors, such as temporary caloric restrictions in relation to a possible preoccupation of the girls with their body weight during adolescence. This may cause similar fluctuations in body weight within a twin pair when the data are not aligned on their maturity status, since both members of a pair are raised in the same family and might influence each other in their dietary intake. In the aligned data, however, a slight difference in the timing of the adolescent growth spurt can cause these fluctuations not to coincide within the pair and thus lower twin correlations, increasing the relative importance of E factors and hence resulting in lower heritability estimates.
being relatively unimportant in explaining stability over the whole age range seen in the small contribution of Et-1 at APHV + 3, does appear to gradually become a more important factor at the older ages in boys (Fig. 3A), which might be construed as reflecting the effects of the gradual accumulation of potentially small differences in, for example, physical activity level or nutrient intake in the twins as they near adulthood. At APHV + 3, summing all transmitted E factors shows that 31.2% of the variation is explained by stable, unique environment (Fig. 3A). In girls, the observed increase in phenotypic tracking for the 3-yr interval APHV to APHV + 3 (r = 0.49; Table 2) vs. the APHV + 1 to APHV + 2 (r = 0.18) appears to be mainly explained by Et, which explains a large amount of the variance on subsequent observations, thus explaining the higher stability in the trait after APHV. At APHV + 2, 42.3% of the variance is explained by E innovation (Ei), which causes instability relative to the previous measurement occasions. A substantial amount of this variance is then subsequently transmitted to the next observation, such that the last 1-yr phenotypic autocorrelation (APHV + 2 to APHV + 3) is markedly higher than the other 1-yr autocorrelations. The total amount of variance explained by Et at APHV + 3 in girls is 44.5%. By summarizing the above, it can be stated that, over the 4-yr interval ranging from 1 yr before APHV to 3 yr after APHV, in both boys and girls, stability of static strength is mainly determined by A factors. When the 2- and 3-yr intervals are considered (second and third subdiagonal of Table 2), the increase in stability over the intervals of the same length, which do not include the time point before APHV, seems to be mediated by both A and E factors in boys and predominantly by E factors in girls. The fact that there still is an increase in stability for intervals of the same length, even after the alignment on age at peak height velocity, possibly is related to the fact that this alignment only takes into account the effect on the variation due to differences in the timing of the adolescent growth spurt and not the effect of differences in tempo of the adolescent growth spurt.

As can be seen in Fig. 3, variance in ARP increases from APHV −1 to APHV + 3 in both boys and girls. This increase in variance during growth is observed for many body dimensions and performance characteristics (23). It may be the consequence of a gradual expression of differences in the genetic potential and individual differences in the accumulation of environmental influences over time or a combination, or interaction, of both. In boys, this increase in variance seems to be due to both a substantial A innovation variance and E innovation variance (Ai and Ei) at each measurement occasion. These variances are subsequently partially transmitted to the subsequent measurement occasions (At and Et). These results suggest that new A factors are expressed at each measurement occasion during male adolescence, with the largest relative increase occurring around APHV (Ai = 30.7%; Table 5), which may be speculated to be related to the rise in testosterone levels and its andabolic effects on muscle mass in male adoles-
ence (23). In girls, the increase in variance is less systematic in terms of Ai and Ei, with a fairly large amount of Ei at APHV + 2 and a fairly large amount of Ai at APHV + 3. Although it can be speculated that motivational factors for a maximal strength effort may introduce some new environmental variance in later adolescence in girls, it is unclear what may be an underlying cause of the amount of Ai at APHV + 3 in girls. It is difficult to compare these findings with those in the literature, since to our knowledge no comparable studies with a comparable amount of repeated measures are available. Katzmarzyk et al. (15) studied the familial aggregation of change scores in a longitudinal study, which also included adolescents; unfortunately, their results cannot be compared with our results due to differences in analytical approach.

In conclusion, the following can be stated. 1) Within the limitations of the statistical power of the present multivariate analysis, A and E sources of variance are shown to be adequate to explain variation in ARP performance during adolescence in both boys and girls. 2) The same A and E factors cause the variation and stability in the trait in boys and girls, yet their absolute and relative contributions to the variation in the trait are not the same. 3) Both A and E sources of variance significantly contribute to the stability in static strength during adolescence. An increase in stability is seen after APHV in both sexes. In boys, this increase can be attributed to both A and E transmitted variance, whereas in girls the increased stability appears to be mainly caused by E transmitted variance.

GRANTS
This research was supported by Research Fund K. U. Leuven (OT/86/80), Nationale Bank van Belgie, Fund for Medical Research (Belgium) (3.0038.82, 3.0008.90, 3.0098.91), and North Atlantic Treaty Organization (860823).

REFERENCES