# Multivariate Genetic Analysis of Lifetime Exercise and Environmental Factors

RIITTA SIMONEN[1], ESKO LEVÄLAHTI[1], JAAKKO KAPRIO[1,2], TAPIO VIDEMAN[3], and MICHELE C. BATTIÉ[4]

*[1]Department of Public Health, University of Helsinki, FINLAND; [2]Department of Mental Health, National Public Health Institute, Helsinki, FINLAND; and [3]Faculty of Rehabilitation Medicine and [4]Department of Physical Therapy, University of Alberta, Edmonton, CANADA*

**ABSTRACT**

SIMONEN, R., E. LEVÄLAHTI, J. KAPRIO, T. VIDEMAN, and M. C. BATTIÉ. Multivariate Genetic Analysis of Lifetime Exercise and Environmental Factors. *Med. Sci. Sports Exerc.*, Vol. 36, No. 9, pp. 1559–1566, 2004. **Purpose:** We investigated whether the association between exercise and individual-specific factors that correlate with exercise may be explained by genetic or common environmental factors. **Methods:** Lifetime exercise data were available from 147 MZ and 153 DZ adult male twin pairs with a mean age of 50 yr (SD = 8 yr). **Results:** The best-fitting quantitative genetic model for adulthood exercise level consisted of additive genetic effects, genetic effects due to dominance and unique environment effects, with genetic effects explaining 51% (95% CI = 29–63%) of the variance. Factors associated with adulthood exercise level were adolescent exercise, participation in competitive sports, perceived health, smoking status, and percent body fat. In bivariate models, approximately half of the covariation between those factors and adulthood exercise level was accounted for by unique environmental effects (i.e., factors not shared by the co-twins). Additive genetic effects explained less (3–20%) than dominance genetic effects (23–53%) of the covariation between those factors and adulthood exercise. Shared environmental effects were present only in the bivariate model of adulthood and adolescent exercise, explaining 11% of the covariance. **Conclusions:** The genetic component shared in common by exercise and factors associated with exercise suggests that there may be a complex pathway of genetic selection and predisposition for a physically active lifestyle. **Key Words:** GENETIC MODEL-FITTING ANALYSIS, TWIN STUDY, PHYSICAL ACTIVITY, SPORTS, LEISURE TIME

Twin studies have shown heritabilities from 39% (10) to 83% (5) of exercise or physical activity levels. There also is a clear pattern of familial aggregation in physical activity levels, indicating that there are genetic and environmental effects shared by family members (1,6,26). Genetic and adolescent environmental effects may have a long lasting impact because physically active adolescents (2,27,29,31), and particularly those participating in competitive sports (2,27,29), tend to be physically active also in adulthood. However, physical activity level fluctuates considerably over time in adolescence and in adulthood, suggesting that environmental factors nevertheless play an important role in the trait.

Variability in physical activity levels has been explained by numerous correlates of physical activity. Physical inactivity has been associated, for example, with smoking (30) and alcohol use (11) and obesity (22). Subjects with good health status tend to have a more physically active lifestyle (2,22). On the other hand, chronic disease may limit physical activity participation. Sociodemographic factors (30) also play role in activity levels. Many of these factors associated with physical activity levels, like physical activity trait, also are at least partly under genetic control. Genetic effects have been found, for example, for smoking (11), alcohol use (12), subjective health status (28), obesity (17), and education level (16).

The underlying causes for the associations of physical activity level and the factors associated with physical activity level remain unknown. It is possible that the genetic influences on physical activity level might be accounted for, at least in part, by genetic influences on those covariates. No study, to our knowledge, to date has attempted to quantify the genetic overlap between physical activity and the correlates of physical activity. If such common biological pathways exist, it would be an important area of information in understanding the predisposition of such phenotypes.

First, the purpose of this study (Fig. 1) was to investigate what proportion of the variability in adulthood exercise level is explained by selected individual-specific environmental factors. Second, the proportion of variability in adulthood physical activity levels and physical activity

Address for correspondence: Riitta Simonen, University of Helsinki, Department of Public Health, P.O. Box 41, 00014 University of Helsinki, Finland; E-mail: riitta.simonen@helsinki.fi.
Submitted for publication September 2003.
Accepted for publication April 2004.

0195-9131/04/3609-1559
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/01.MSS.0000074670.03001.98

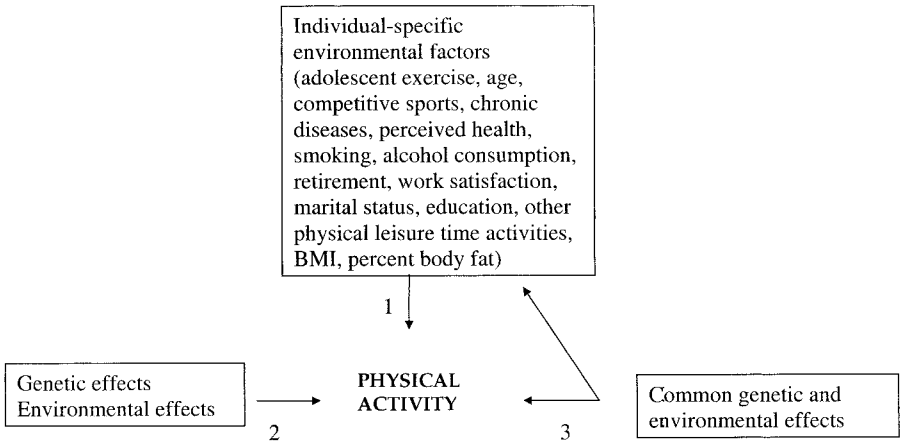


**FIGURE 1—Aims of the study. First, the factors associated with exercise were determined. Second, the proportion of variability ascribable to genetic and environmental effects was estimated. Third, common genetic or environmental effects between physical activity and the individual-specific factors were studied.**

levels in adolescence ascribable to additive and dominance genetic effects, as well as for shared and unique environmental effects, was estimated using a quantitative genetic model among MZ and DZ male twin pairs. Third, we explored whether the association between such individual-specific factors and exercise was determined by genetic or environmental effects in common.

## METHODS

**Subjects.** Subjects for this Twin Spine Study were selected from the population-based Finnish Twin Cohort consisting of all twins born before 1958 who were alive in 1975. The selection was based on discordance between co-twins for a specific common behavioral or environmental factor (exercise, smoking as well as occupational sitting, lifting, and driving), described in more detail elsewhere (4). Those selection criteria were based on the availability of relevant prior information from the Finnish Twin Cohort 1975 and 1981 questionnaires and the suspected importance in the etiology of spinal degeneration and back symptom complaints, which was the primary study goal of the Twin Spine Study. In addition, the subjects included 33 randomly selected MZ and DZ pairs. A total of 147 MZ and 153 DZ male twin pairs (aged 35–70 yr) were studied. The validity of questionnaire-based diagnosis of zygosity was studied previously in a subsample of 104 twin pairs. There was 100% agreement in classification between the questionnaire data and 11 blood markers, with an estimated probability of misclassification of 1.7% at the population level (25).

**Data acquisition.** A structured interview, lasting 2.5 h on average, was used to obtain data on health status and health habits. Lifetime exercise participation from age 12 through the present took from a few minutes to about an hour, depending on the exercise activity level. Information for every event lasting at least 3 months was collected for exercise and other physical leisure time activities than exercise. This information included the time span of participation in years, months per year of participation, and mean frequency (times per week), duration (minutes per session), and intensity (light, moderate, strenuous), and whether participation was at a competitive level if sports related (yes/no). The summary score was calculated by summing the weekly hours separately for exercise and other physical leisure time activities both for adolescence (age 12–18) and for adulthood (age 18 to age at interview). The 5-yr interval repeatability ($N = 38$) of the lifetime exercise interview yielded acceptable repeatability (intraclass correlation: 0.73) for the exercise parameter (lifetime mean hours per week) (24). Friedenreich et al. (9) have used similar method for lifetime exercise interview among breast cancer patients, and they also got very similar repeatability (0.72) for exercise and sports activities.

Demographic information and health history, chronic diseases, perceived health compared with people of same age were elicited by interview and lifetime and current smoking, in terms of cigarettes smoked daily and years smoked as well as current alcohol consumption were obtained by interview, as well. Smoking was further converted into pack years and alcohol use into grams of absolute alcohol per month. Interviewers were blind with respect to the specific discordance or selection criteria. Clinical examination included basic anthropometric measures, such as weight, height, and percent body fat by bioimpedance measurements (Spectrum II, RJL Systems, Detroit, MI). Bioelectrical impedance for the co-twins was measured at the same time by trained personnel. Subjects did not drink coffee or eat meal for 2 h before the measurements. Metal objects, such as wristwatches and purses containing coins were taken off. Subjects were measured in the supine position, with hands and legs a little apart from the body. Electrodes were placed on the ipsilateral side of the body of the dominant hand. The electrode on the hand was placed on the wrist near the ulnar head. Another electrode was placed just behind the middle finger. The electrode on the foot was placed near the middle malleolus in the mid-line of the ankle while another electrode was placed behind the middle toe. In healthy adults, the bioelectrical impedance intertester and interday variability has been 0.02 kg fat free mass (95% confidence interval [CI] = $-0.2 \pm 0.2$) with a coefficient of variation of 1.3% (15). The study protocols were reviewed and approved by the Ethical Committee of the Department of Public Health at the University of Helsinki. Informed written consent was obtained from all participants.

**Statistics.** In statistical analysis, adulthood and adolescent exercise, other leisure time activities, number of chronic diseases, alcohol consumption, and smoking were categorized because of skewness. Observations of percent body fat and body mass index were transformed for analyses.

The equality of means of transformed variables by zygosity was tested by an adjusted Wald test to take into account that the twin individuals had been sampled as twin pairs and thus do not represent fully independent observations. The equality of variances in MZ and DZ twins were tested using variance ratio test. The equality of distributions of categorical variables by zygosity was tested by design based independence test. Wilcoxon rank-sum test (also known as Mann-Whitney test) was used to determine whether the age distributions of MZ and DZ twin pairs were the same.

The effect of suspected determinants and adolescent exercise on adulthood exercise was modeled using survey logistic regression (LR) and survey ordered logistic regression (OLR), which takes into account the covariance within twin pair and calculates 95% CI of estimates adjusted to that within pair covariance.

LR was used to get estimates of odd ratios (OR) for each of the exercise participating categories compared with inactive exercise categories.

OLR was used to estimate the linear effect of adolescent exercise and determinants on adulthood exercise. In this method, the change of one unit in adolescent exercise or determinant was assumed to cause the same change for every cutpoint of adulthood exercise categories. All factors significant in univariate models were put in the multivariate OLR model. STATA SVY-procedures were used to carry out all these analyses.

Univariate variance component models for exercise variables were fitted on contingency tables using maximum likelihood estimation. In multivariate analysis, variables were summarized to polychoric correlations, asymptotic variances, and asymptotic covariances. Asymptotic weighted least squares estimation method was used. All estimation was done in Mx, a statistical program designed for analyses of genetically informative data.

The classical twin study design assumes that genetic effects and environmental effects are not correlated, there is no genotype-environment interaction, and there is random mating with respect to the traits under study (20). It also assumes that MZ and DZ twins come from the same base population; hence, they should not differ in means or variances. If these assumptions hold true, the twin model can be used to determine the relative contribution of additive effect of genes (A), dominance effect of genes (D), common environment (C), and unique environment (E). MZ pairs share genetic effects fully but also have the same common environment. In contrast, DZ pairs share half of their genes by descent in common but have the same common environment. Thus, the increased similarity of both MZ and DZ twins can be used to estimate common environmental effects, whereas the greater similarity of MZ pairs compared with DZ pairs provides evidence for genetic effects. The extent to which MZ pairs are more similar than DZ pairs permits approximation of additive from dominance effects (20). First, univariate path models were estimated. This represents the potential sources of variance that can be estimated in a twin model, but in univariate cases, effect due to dominance and common environment cannot be simultaneously estimated. The squares of the standardized path coefficients (a, c, d, e) from A, C, D, and E, respectively, are estimates of the variance components.

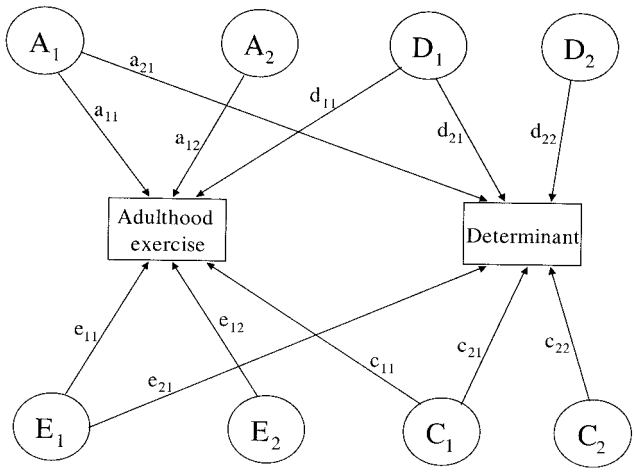


**FIGURE 2—Bivariate genetic factor model. The determinants indicate adolescent exercise, age, competitive sports, chronic diseases, perceived health, smoking, alcohol consumption, retirement, work satisfaction, marital status, education, other physical leisure time activities, BMI, and percent body fat.**

All determinants with statistically significant linear effects on adulthood exercise in multivariate OLR–analysis were used in bivariate twin analysis. Cholesky decomposition genetic factor model was used to analyze the genetic and environmental relationships between adulthood exercise and determinants as well as adolescent exercise (Fig. 2). The most parsimonious model was selected based on Akaike's information criterion (AIC) (3) and biological relevance. Proportions of variance explained by genetic and environmental factors and the genetic and environmental correlations of the most parsimonious models are reported here. A higher proportion of variance indicates that the same genes or environmental effects influence more than one trait at a time, whereas a lower proportion of variance indicates an influence of different genes or environmental effects. The contribution of genes or environmental effects to the observed phenotypic correlation between two variables is measured by genetic or environmental correlation (20).

## RESULTS

**Descriptive statistics and determinants for exercise.** Table 1 shows the distributions of studied determinants and the exercise levels in MZ and DZ twins, as well as results of testing for differences in means and distributions. The numbers of chronic disease detected by a physician in this cohort were cardiovascular disease (147),

TABLE 1. Distribution of exercise and suspected determinants of adulthood exercise among MZ and DZ twin individuals.

| Variable | Category | MZ Twins ($N$ = 147 pairs) | DZ Twins ($N$ = 153 pairs) | Test for Equality of Distribution by Zygosity ($P$) |
|---|---|---|---|---|
| Adulthood exercise (from age 18 to present) | 0–<0.1 h·wk$^{-1}$ | 37 (13) | 38 (12) | |
| | 0.1–<1.1 h·wk$^{-1}$ | 71 (24) | 58 (19) | 0.25 |
| | 1.1–<3 h·wk$^{-1}$ | 78 (27) | 103 (34) | |
| | 3–<6 h·wk$^{-1}$ | 62 (21) | 70 (23) | |
| | 6+ h·wk$^{-1}$ | 46 (16) | 37 (12) | |
| Adolescent exercise (age 12 to 18) | 0–<0.1 h·wk$^{-1}$ | 61 (21) | 59 (19) | |
| | 0.1–<1.1 h·wk$^{-1}$ | 36 (12) | 43 (14) | 0.97 |
| | 1.1–<3 h·wk$^{-1}$ | 75 (26) | 78 (25) | |
| | 3–<6 h·wk$^{-1}$ | 56 (19) | 59 (19) | |
| | 6+ h·wk$^{-1}$ | 66 (22) | 67 (22) | |
| Competitive sports | No | 164 (56) | 172 (56) | 0.93 |
| | Yes | 130 (44) | 134 (44) | |
| Chronic diseases | No | 130 (45) | 169 (55) | |
| | Yes | 162 (55) | 137 (45) | <0.05 |
| Perceived health | Much better | 28 (10) | 19 (6) | |
| | Somewhat better | 104 (35) | 97 (32) | |
| | Same | 120 (41) | 157 (51) | |
| | Slightly worse | 36 (12) | 30 (10) | 0.10 |
| | Much worse | 6 (2) | 3 (1) | |
| Smoking | Never-smoker | 103 (35) | 111 (36) | |
| | <25 pack yr | 140 (48) | 141 (46) | |
| | 25–80 pack yr | 51 (17) | 54 (18) | 0.94 |
| Alcohol consumption | 0 g·month$^{-1}$ | 28 (10) | 25 (8) | |
| | <100 g·month$^{-1}$ | 64 (22) | 77 (25) | |
| | <430 g·month$^{-1}$ | 109 (37) | 116 (38) | |
| | <680 g·month$^{-1}$ | 64 (22) | 55 (18) | 0.75 |
| | 680–3748 g·month$^{-1}$ | 29 (10) | 33 (11) | |
| Retired | No | 247 (84) | 257 (84) | |
| | Yes | 47 (16) | 49 (16) | 0.99 |
| Work satisfaction | Almost always | 156 (54) | 148 (49) | |
| | Not so often–almost never | 132 (46) | 156 (51) | 0.22 |
| Marital status | Married/divorced/widowed | 267 (91) | 270 (88) | 0.35 |
| | Single | 27 (9) | 36 (12) | |
| Education level | <9 yr | 221 (75) | 210 (69) | |
| | 9 yr or more | 73 (25) | 96 (31) | 0.16 |
| Other leisure time activities | 0–<0.1 h·wk$^{-1}$ | 42 (14) | 45 (15) | |
| | 0.1–<2 h·wk$^{-1}$ | 121 (40) | 118 (39) | |
| | 2–<12 h·wk$^{-1}$ | 124 (41) | 112 (37) | 0.15 |
| | 12+ h·wk$^{-1}$ | 13 (4) | 25 (8) | |
| BMI†§ | | 25.9 (25.4–26.4) | 25.8 (25.4–26.2) | 0.76 |
| Fat%†# | | 23.9 (23.0–24.7) | 21.1 (20.5–21.7) | <0.01 |

† Mean (95% confidence interval) and test of equality of means.
§ MZ, $N$ = 147 pairs; DZ, $N$ = 152 pairs.
# Test for equality of SD: $P < 0.01$; MZ: $N$ = 137 pairs; DZ: $N$ = 150 pairs.

musculoskeletal disease (117), respiratory disease (84), endocrine and metabolic system disease (53), mental disease (27), and neurological disease (21). The proportion with chronic diseases was higher among MZ twins than DZ twins ($P < 0.05$). Also, both mean ($P < 0.001$) and variance ($P < 0.01$) of percent body fat were higher for MZ twins. There were no other significant differences in MZ and DZ distributions.

In OLR models (Table 2), significant correlates of adulthood exercise were adolescent exercise, competitive sports in adolescence, perceived health, smoking, and percent body fat ($P < 0.001$ except for percent body fat $P = 0.02$). Also most OR from LR analysis (Table 2) were significant for factors that were significant in the OLR analysis. Marital status (single vs others) and education level also had two significant OR. However, there was no trend in OR in exercise levels with marital status and education, but there were significantly increased probabilities for the single and better educated to belong to groups that exercise 1.1–3 h·wk$^{-1}$ and to 3–6 h·wk$^{-1}$ compared with belonging to the inactive group.

**Polychoric twin correlations.** Polychoric twin correlation of MZ pairs for adulthood exercise was substantially higher (r = 0.544, 95% CI = 0.413–0.650) than the correlations observed among DZ pairs (r = 0.133, 95% CI = 0.0–0.306), which suggests a contribution of genetic effects, including effects due to dominance. However, in adolescent exercise the correlation of MZ (r = 0.543, 95% CI = 0.40–0.657) versus DZ (r = 0.455, 95% CI = 0.297–0.586) suggested that environmental effects shared by the pairs could be expected to contribute to the observed variance in this phenotype.

**Univariate model fitting.** The genetic and environmental variance components for adulthood and adolescent exercise levels were estimated by standard univariate twin analysis. Only the model with unique environment (E) alone could be rejected. The model resulting in the lowest AIC (−55.70, $\chi^2$ = 28.30) for the adulthood exercise was a

TABLE 2. Individual specific factors associated with adulthood exercise level, odd ratios (OR) using survey logistic regression models and coefficients and OR estimated using survey ordered logistic regression are presented.

| | OR for Each Exercise Level (Compared with Inactive) | | | | Ordered Logistic Regression | |
|---|---|---|---|---|---|---|
| Individual-Specific Factors | 0.1–<1.1 | 1.1–<3 | 3–<6 | 6– | Univariate OR (95% CI) | Multivariate OR (95% CI) |
| Adolescent exercise | 1.67 (1.27–2.19) | 2.08 (1.60–2.71) | 2.43 (1.89–3.12) | 2.97 (2.17–4.07) | 1.96 (1.70–2.25) | 1.76 (1.51–2.04) |
| Age | 0.76 (0.52–1.09) | 1.03 (0.68–1.56) | 0.90 (0.60–1.34) | 0.77 (0.49–1.21) | 0.97 (0.78–1.22) | — |
| Competitive sports | 3.90 (1.73–8.82) | 7.25 (3.30–15.9) | 10.1 (4.29–23.5) | 20.6 (8.14–52.1) | 3.54 (2.53–4.94) | 1.89 (1.30–2.74) |
| Chronic diseases ($N$ = 598) | 0.88 (0.49–1.60) | 0.95 (0.55–1.67) | 0.79 (0.43–1.45) | 1.10 (0.571–2.11) | 1.00 (0.74–1.35) | — |
| Perceived health | 1.18 (0.82–1.70) | 1.70 (1.19–2.44) | 1.95 (1.37–2.79) | 2.18 (1.47–3.23) | 1.62 (1.35–1.96) | 1.35 (1.11–1.64) |
| Smoking | 0.73 (0.49–1.08) | 0.49 (0.32–0.73) | 0.49 (0.32–0.76) | 0.41 (0.25–0.66) | 0.65 (0.52–0.81) | 0.74 (0.58–0.92) |
| Alcohol consumption | 1.14 (0.87–1.49) | 0.99 (0.75–1.30) | 1.04 (0.80–1.36) | 1.04 (0.76–1.41) | 0.98 (0.85–1.13) | — |
| Retirement | 0.68 (0.30–1.54) | 0.76 (0.37–1.56) | 0.51 (0.23–1.11) | 0.68 (0.28–1.65) | 0.79 (0.51–1.22) | — |
| Work satisfaction* | 1.07 (0.59–1.93) | 1.01 (0.58–1.75) | 0.75 (0.41–1.35) | 0.78 (0.40–1.52) | 0.80 (0.60–1.08) | — |
| Marital status (singles vs others) | 0.49 (0.21–1.14) | 0.26 (0.11–0.61) | 0.4 (0.18–0.91) | 0.61 (0.26–1.42) | 0.71 (0.39–1.30) | — |
| Marital status (divorced/widowed vs others) | 1.38 (0.54–3.50) | 1.34 (0.54–3.36) | 1.15 (0.44–3.02) | 1.04 (0.36–3.00) | 0.94 (0.61–1.45) | — |
| Education level | 1.27 (0.64–2.49) | 1.98 (1.01–3.90) | 2.07 (1.05–4.09) | 1.11 (0.49–2.52) | 1.20 (0.88–1.65) | — |
| Other leisure time activities | 1.29 (0.89–1.87) | 1.54 (1.04–2.29) | 1.29 (0.90–1.86) | 1.36 (0.92–2.00) | 1.14 (0.93–1.40) | — |
| BMI** | 1.40 (0.54–3.62) | 1.73 (0.67–4.50) | 1.13 (0.44–2.87) | 0.68 (0.25–1.81) | 0.76 (0.46–1.27) | — |
| Percent body fat*** | 0.93 (0.53–1.62) | 0.76 (0.43–1.37) | 0.66 (0.36–1.19) | 0.57 (0.31–1.05) | 0.69 (0.50–0.94) | 0.69 (0.51–0.94) |

* $N$ = 592; ** $N$ = 598; *** $N$ = 574.

model with additive genetic (A), dominance genetic (D), and unique environment (E) effects. However, it is unlikely that the ADE model would be the most parsimonious because the estimate of the additive genetic component was zero. Likewise, the ACE model included an estimate of zero for shared environment suggesting that it also is not the best model. The next best model, in terms of the second lowest AIC, was an AE model with a combination of additive genetic (51%, 95% CI = 38–63%) and unique environment (49%, 95% CI = 37–62%) yielding an AIC of −55.51 ($\chi^2$ = 30.49) and thus being the most parsimonious and biologically relevant model in explaining adulthood exercise level.

Similarly, the model E was rejected also for adolescent exercise. In the ADE model, the estimate of the dominance effects was zero, thus making this an unlikely model. The lowest AIC for this phenotype was with a CE model (AIC = −51,41, $\chi^2$ = 34.59), which includes common environment and unique environment, both explaining 50% of the total variance. The next best model with a marginally worse fit ($\Delta\chi^2$ = 0.8) for adolescent exercise was a model with A-, C-, and E-components (AIC = −50.25, $\chi^2$ = 33.76), with additive genetic effects explaining 18% (95% CI = 0–60%), shared environment 37% (95% CI = 3–58%), and unique environment 46% (95% CI = 34–59%) of the total variance in adolescent exercise. The model AE consisting of additive genetic (58%, 95% CI = 46–68%) and unique environment (42%, 95% CI = 32–54%) provided slightly worse fit than the ACE model.

**Bivariate genetic analysis.** The individual-specific factors that were associated with adulthood exercise level in the multivariate OLR analyses were put in a series of bivariate genetic models with exercise level to investigate the genetic and environmental correlations between the exercise and other phenotypes (Table 3). The bivariate models of exercise and BMI as well as education level did not reach good model fit and were therefore omitted from further modeling. Adulthood exercise had additive genetic variation in common with adolescent exercise, competitive sports, perceived health, smoking, and percent body fat. For percent body fat and smoking, the genetic correlation was −1, suggesting that additive genetic effects caused higher percent body fat and that smoking has effects that decrease exercise level. Otherwise correlations were unity suggesting that additive genetic effects on greater adolescent exercise level, participation in competitive sports and better perceived health are also contributing to increased adulthood exercise level. Also, the shared environment of twins con-

TABLE 3. Most parsimonious bivariate genetic factor models for adulthood exercise: cross-twin cross-trait correlations, additive genetic correlation($r_a$), dominance genetic correlation ($r_d$), common environmental correlation ($r_c$), and unique environmental correlation ($r_c$) and fit statistics; 95% confidence interval is expressed in parentheses.

| | Cross-Twin Cross-Trait Correlations¶ with Adulthood Exercise | | Correlations with Adulthood Exercise | | | | Fit Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Determinant | MZ | DZ | A | C | E | Proportion of unique environmental variation in adulthood exercise accounted for by correlation with unique environment in determinant | $\chi^2$ | df | $P$ | AIC |
| Adolescent exercise | 0.41 | 0.28 | 1* | 1* | 0.26 (0.07–0.44) | 0.07 | 8.07 | 6 | 0.23 | −3.93 |
| Competitive sports | 0.44 | 0.06 | 1* | — | 0.48 (0.18–0.78) | 0.24 | 10.96 | 8 | 0.20 | −5.04 |
| Perceived health | 0.17 | 0.05 | 1* | — | — | — | 7.70 | 9 | 0.57 | −10.30 |
| Smoking | −0.18 | 0.01 | −1* | — | — | — | 8.07 | 9 | 0.53 | −9.93 |
| Fat%† | −0.15 | 0.04 | −1* | — | — | — | 3.96 | 9 | 0.91 | −14.04 |

¶ Estimated using double-entry.
* Fixed: single variance component is accounting for all additive genetic variance/common environmental variance.
† All missing values in fat% imputed.

TABLE 4. Most parsimonious bivariate genetic factor models for adulthood exercise: proportions (and 95% confidence intervals) of adulthood exercise and determinants explained by additive genetic factors (A), dominance genetic factors (D), common environmental factors (C), and unique environmental factors (E) (see Fig. 1).

| | Adulthood Exercise: Proportions of Variance Components | | | | Determinant: Proportions of Variance Components | | |
|---|---|---|---|---|---|---|---|
| Determinant | $A_1$, $A_2$ | $D_1$ | $C_1$, $C_2$ | $E_1$ | $A_1$ | $C_1$ | $E_1$, $E_2$ |
| Adolescent exercise | 0.20 (0.00–0.55) | 0.23 (0.00–0.37) | 0.11 (0.00–0.31) | 0.46 (0.35–0.58) | 0.24 (0.00–0.60) | 0.35 (0.06–0.60) | 0.41 (0.28–0.53) |
| Competitive sports | 0.13 (0.05–0.24) | 0.42 (0.25–0.54) | — | 0.45 (0.33–0.58) | 0.81 (0.71–0.92) | — | 0.19 (0.08–0.29) |
| Perceived health | 0.13 (0.05–0.28) | 0.43 (0.26–0.57) | — | 0.44 (0.32–0.55) | 0.34 (0.20–0.48) | — | 0.66 (0.52–0.80) |
| Smoking | 0.05 (0.01–0.13) | 0.52 (0.38–0.64) | — | 0.43 (0.31–0.55) | 0.52 (0.36–0.67) | — | 0.48 (0.33–0.64) |
| Fat% | 0.03 (0.00–0.07) | 0.53 (0.41–0.65) | — | 0.44 (0.32–0.57) | 0.61 (0.45–0.77) | — | 0.39 (0.23–0.55) |

tributing to adolescent exercise had an effect on adulthood exercise level. Unique environment of adolescent and adulthood exercise changed comparatively more than shared environment and additive genes. There the unique environment correlation ($r_e$) was 0.26, indicating that 7% of unique environmental variation in adulthood exercise was accounted for unique environmental variation in adolescent exercise. Similarly, 24% of the unique environmental variation of adulthood exercise was accounted for by the unique environment of competitive sports (unique environment correlation $r_e = 0.48$).

Although adolescent exercise, competitive sports, perceived health, smoking, and percent body fat each had an additive genetic correlation of 1 or −1 with exercise level, there are also dominance effects on adulthood exercise level (see Table 4; $D_1$ variance components for exercise) not affecting adolescent exercise, competitive sports, perceived health, smoking, and percent body fat (see Table 4; $D_1D_2$ variance components for determinants). All bivariate models for exercise and the studied factors suggest that proportion of combined additive and dominance genetic effects is 0.42–0.57 and estimates for the proportion of dominance genetic effects is 0.23–0.53. Most models suggest that no shared environmental factors contribute to adulthood exercise level except the model with adolescent exercise, which suggests a significant proportion of shared environmental effects (0.11, 95% CI = 0.00–0.31).

## DISCUSSION

In this cohort of Finnish adult male twins, one-half of the variation in adulthood exercise level and one-sixth of the variability in adolescent exercise level were accounted for by genetic variation. Factors associated with adulthood exercise level were adolescent exercise, participation in competitive sports in adolescence, perceived health, smoking status, and percent body fat. Those factors also shared additive and dominance genetic components in common with adulthood exercise.

In the univariate model for adulthood exercise, the parameter estimates for the model consisting of additive genetic, dominance genetic, and unique environment (ADE) and the model of additive genetic, common environment, and unique environment (ACE) were close to the model of additive genetic and unique environment (AE). However, in the model of additive genetic, dominance genetic, and unique environment (ADE), the estimate of the additive genetic component was zero as was the common environment component in the model of additive genetic, common environment, and unique environment (ACE), suggesting that these models were unlikely to be the most plausible ones for adulthood exercise. Therefore, the most parsimonious model in explaining adulthood exercise level is one of additive genetic and unique environment (AE), which is similar to the findings of Koopmans et al. (13) and Frederiksen and Christensen (8). Further modeling showed that there were also dominance effects among all covariations as well as shared environment effects for the covariation of adolescent and adulthood exercise.

The heritability estimate of 51% for adulthood exercise in this study is well within the mid range of the exercise heritability estimates of previous twin studies, which have shown heritabilities of 39% (10), 45% (13), 49% (8), 53% (14), 62% (11), 68% (18), and 83% (5). Conversely, in several family studies, where parent and offspring physical activity levels have been investigated, either nonsignificant (1,6) or low (26) heritability estimates have been found. It should be kept in mind that heritability is a population-specific parameter (23), which might in part explain the differences in the heritability estimates between this study and the previous studies using a variety of populations. In addition, unequal gender effects can be one source of underestimation of heritability in family studies. Several studies (1,13) suggest that familial resemblance occurs within generations (spouses and siblings) rather than across generations (parent-offspring) indicating either a cohort effect or age-specificity for this phenotype. When a segregation analysis revealing potential familial gene transmission patterns is not used, the effect of genetic and shared family environmental factors on the trait cannot be distinguished in family studies. Also of importance, a relatively high (correlation = 0.22–0.43) spouse resemblance in physical activity (26) can be indicative of shared environmental effects during the year of cohabitation, nonrandom (assortative) mating or social homogamy. In twin studies, the high spouse correlation may underestimate the heritability coefficient by raising the DZ correlation (21).

There was an interesting finding of increasing importance of genetics on exercise level from adolescence (17%) to adulthood (51%). This may indicate that there may be more

freedom for voluntary choices in adulthood lifestyle than in adolescence, with the choices that may be made based on inherited predispositions in exercise performance. However, a longitudinal research design is needed to show the trend of inheritance more reliably across the lifespan.

There is evidence that physically more active subjects tend to have better subjective health (2), be less likely to smoke (13,19), and have lower risk for obesity (19). Compared with less physically active adults, more active adults have been more likely to have exercised during adolescence (3,7,29,31), and their exercise often has had a competitive nature (29). Some of these associations between physical activity and its determinants were complex. For example, no linear association with physical activity was detected in education level and marital status. In education level, greater education level indicated greater physical activity level from 1 h up to 6 h of exercise per week. However, this association did not exist among the least and most active subjects. A similar trend seemed to be seen in marital status, where subjects exercising from 1 to 6 $h \cdot wk^{-1}$ were less likely to be single than the least and most active subjects. In this study, modeling of the phenotypic association between adulthood exercise and the factors associated with exercise level revealed that about half of the phenotypic covariation was due to unique environmental effects, the rest being shared environmental or genetic effects. The bivariate analysis indicated that there were overlapping genetic influences between adulthood exercise and adolescent exercise (43%), as well as adulthood exercise and competitive sports (55%). This finding suggests that there may be a general biological explanation for lifelong exercise participation. In terms of competitive sports, the exercise intensity and greater frequency in exercise involvement may facilitate the exercise lifestyle. Similarly, genetic factors in common explain over a half of the covariation between adulthood exercise and smoking (57%), adulthood exercise, and perceived health (56%), and adulthood exercise and percent body fat (56%).

Finally, results reported in this study need to be interpreted in the context of some limitations. First, the fact that the twins are all men limits the extent to which these results are generalizable to women. For example, genetic effects on physical activity levels were considerably smaller for women than for men in a Portuguese twin cohort (17) and in a Belgium twin cohort (6). Second, although this lifetime exercise interview yielded a good reliability coefficient (ICC = 0.73) with a 5-yr test-retest interval, participation in competitive sports was less reliable (ICC = 0.58), and the repeatability of the adolescent exercise level variable has not been assessed separately (24). It may be assumed that exercise events that are remote in time could be either poorly remembered or even overestimated by the most physically active subjects. For adulthood exercise level, power calculations (21) revealed that when there is no common environmental component and the additive component explains 51% of the variation of the phenotype, the power is sufficient (power value = 1.00) to detect such an additive genetic component. However, for adolescent exercise, where the common environment explains 37% of the phenotypic variation, the power is less than 0.40 indicating weak statistical power to detect the 18% additive genetic component in a population of this sample size. The lack of power may thus hamper the generalization of the results concerning exercise in adolescence in this study.

The selection criteria of this cohort from the population based Finnish Twin Cohort may raise some concern. Due to the discordance in physical activity level, the variance in physical activity may have increased, but the sample contained also twins chosen at random and discordant for other reasons. However, these selection criteria were applied both for MZ and DZ twin pairs. Furthermore, the overall profile in exercise level in this cohort is quite similar to the exercise level in the cohort of origin. Also, the estimates of genetic variance in the covariates are close to those obtained in other studies. Therefore, we consider it unlikely that the selection process produced major biases in the relationships between physical activity and covariates that are a focus of the genetic models.

In line with many of the previous twin studies, we found that half of the variability in adulthood exercise level was accounted for by genetic effects. Several individual-specific factors that were correlates of exercise had a genetic component in common with exercise: adolescent exercise, competitive sports in adolescence, perceived health, smoking, and percent body fat. The contribution of environmental factors to adulthood exercise, however, was notable ranging from 43 to 46% of the variation in the bivariate models. This study suggests that successful modification of environmental factors may have a potential effect in increasing exercise levels in adulthood. The genetic overlap suggests possible underlying biologic mechanism for the observed phenotypic associations between the traits, which may ultimately help in the identification of specific genes involved in exercise participation.

The authors acknowledge the financial support of the Finnish Academy, Ministry of Education, and the European Union funding for the project EURODISC (QLK6-CT-2002-02582).

## REFERENCES

1. Aarnio, M., T. Winter, U. M. Kujala, and J. Kaprio. Familial aggregation of leisure-time physical activity: a three generation study. *Int. J. Sports Med.* 18:549–556, 1997.
2. Barnekow-Bergkvist, M., G. Hedberg, U. Janlert, and E. Jansson. Physical activity pattern in men and women at the ages 16 and 34 and development of physical activity from adolescence to adulthood. *Scand. J. Med. Sci. Sports* 6:359–370, 1996.
3. Barnekow-Bergkvist, M., G. Hedberg, U. Janlert, and E. Jansson. Prediction of physical fitness and physical activity level in adulthood by physical performance and physical activity in adolescence: an 18-year follow-up study. *Scand. J. Med. Sci. Sports* 8:299–308, 1998.
4. Battié, M. C., T. Videman, L. E. Gibbons, L. D. Fisher, H. Manninen, and K. Gill. Volvo award in clinical sciences: determinants of lumbar disc degeneration. A study relating lifetime

exposures and magnetic resonance imaging findings in identical twins. *Spine* 20:2601–2612, 1995.
5. Beunen, G., and M. Thomis. Genetic determinants of sports participation and daily physical activity. *Int. J. Obes.* 23:S55–S63, 1999.
6. Boomsma, D. I., M. B. M. van den Bree, J. F. Orlebeke, and C. M. Molenaar. Resemblances of parents and twins in sports participation and heart rate. *Behav. Genet.* 1:123–141, 1989.
7. Dennison, B. A., J. H. Straus, E. D. Mellits, and E. Charney. Childhood physical fitness tests: predictor of adult physical activity levels? *Pediatrics* 82:324–330, 1988.
8. Frederiksen, H., and K. Christensen. The influence of genetic factors on physical functioning and exercise in second half of life. *Scand. J. Med. Sci. Sports. Exerc.* 13:9–18, 2003.
9. Friedenreich, C. M., K. S. Courneya, and H. E. Bryant. The lifetime total physical activity questionnaire: development and reliability. *Med. Sci. Sports Exerc.* 30:266–274, 1998.
10. Heller, R. F., D. L. O'Connell, C. K. Roberts, et al. Lifestyle factors in monozygotic and dizygotic twins. *Genetic Epidemiol.* 5:311–321, 1988.
11. Kaprio, J., M. Koskenvuo, and S. Sarna. Cigarette smoking, use of alcohol, and leisure-time physical activity among same-sexed adult male twins. In: *Twin Research 3: Epidemiological and Clinical Studies.* New York: Alan R. Liss, Inc, 1981, pp. 37–46.
12. Kaprio, J., M. Koskenvuo, H. Langinvainio, K. Romanov, S. Sarna, and R. J. Rose. Genetic influences on use and abuse of alcohol: a study of 5638 adult Finnish twin brothers. *Alcohol Clin. Exp. Res.* 11:349–356, 1987.
13. Koopmans, J. R., L. J. P. van Doornen, and D. I. Boosma. Smoking and sports participation. In: *Genetic Factors in Coronary Heart Disease*, U. Goldbourt, U. de Faire, and K. Berg (Eds.). Dordrecht, The Netherlands: Kluwer Academic, 1994, pp. 217–235.
14. Kyle, U. G., L. Genton, D. O. Slosman, and C. Pichard. Fat-free and fat mass percentiles in 5225 healthy subjects aged 15 to 98 years. *Nutrition* 17:534–541, 2001.
15. Lauderdale, D. S., R. Fabsitz, J. M. Meyer, P. Sholisnky, V. Ramakrishnan, and J. Goldberg. Familial determinants of moderate and intense physical activity: a twin study. *Med. Sci. Sports Exerc.* 29:1062–1068, 1997.
16. Liechtenstein, P., N. Pedersen, and G. E. McClearn. The origins of individual differences in occupational status and education level. *Acta Sociologica* 35:13–31, 1992.
17. Maes, H. H., M. C. Neale, and L. J. Eaves. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* 27:325–351, 1997.
18. Maia, J. A., M. Thomis, and G. Beunen. Genetic factors in physical activity levels: a twin study. *Am. J. Prev. Med.* 23(2 Suppl. 1):87–91, 2002.
19. Marti, B., J. Pekkanen, A. Nissinen, et al. Association of physical activity with coronary risk factors and physical ability: twenty-year follow-up of a cohort of Finnish men. *Age Ageing* 18:103–109, 1989.
20. Neale, M. C., and L. R. Cardon. *Methodology for Genetic Studies of Twins and Families*. Dordrecht, The Netherlands: Kluwer Academic, 1992, pp. 149–259.
21. Neale, M. C., L. J. Eaves, and K. S. Kendler. The power of the classical twin study to resolve variation in threshold traits. *Behav. Genet.* 24:239–258, 1994.
22. Norman, A., R. Bellocco, F. Vaida, and A. Wolk. Total physical activity in relation to age, body mass, health and other factors in a cohort of Swedish men. *Int. J. Obes.* 26:670–675, 2002.
23. Plomin, R., J. C. Defries, G. E. McClearn, and P. McGuffin. *Behavioral Genetics*, 4th Ed. New York: Worth Publishers, 2000, pp. 85–92.
24. Ropponen, A., E. Levälahti, R. Simonen, T. Videman, and M. C. Battié. Repeatability of life-time exercise reporting. *Scand. J. Med. Sci. Sports* 11:185–192, 2001.
25. Sarna, S., J. Kaprio, P. Sistonen, and M. Koskenvuo. Diagnosis of twin zygosity by mailed questionnaire. *Hum. Hered.* 28:241–254, 1978.
26. Simonen, R. L., L. Pérusse, T. Rankinen, T. Rice, D. C. Rao, and C. Bouchard. Familial aggregation of physical activity levels: the Québec Family Study. *Med. Sci. Sports Exerc.* 34:1137–1142, 2002.
27. Simonen, R. L., T. Videman, J. Kaprio, E. Levälahti, and M. C. Battié. Factors associated with exercise lifestyle: a study of monozygotic twins. *Int. J. Sports Med.* 24:499–505, 2003.
28. Svedberg, P., P. Lichtenstein, and N. L. Pedersen. Age and sex differences in genetic and environmental factors for self-rated health: a twin study. *J. Gerontol. B. Psychol. Sci. Soc. Sci.* 56: S171–S178, 2001.
29. Telama, R., L. Laakso, X. Yang, and J. Viikari. Physical activity in childhood and adolescence as predictor of physical activity in young adulthood. *Am. J. Prev. Med.* 13:317–323, 1997.
30. Trost, S. G., N. Owen, A. E. Bauman, and J. F. Sallis. Correlates of adults' participation in physical activity: review and update. *Med. Sci. Sports Exerc.* 34:1996–2001, 2002.
31. Vanreusel, B., R. Renson, G. Beunen, et al. A longitudinal study of youth sport participation and adherence to sport in adulthood. *Int. Rev. Soc. Sport.* 32:373–387, 1997.