Genetic Factors in Physical Activity Levels
A Twin Study

José A. R. Maia, PhD, Martine Thomis, PhD, Gaston Beunen, PhD

Background: Substantial interindividual variation is observed in sports participation and physical activity levels in youth. This study aimed to (1) estimate the relative contribution of genes, along with shared and nonshared environmental factors, to variation in sports participation index (SPI) and leisure-time physical activity (LTPA); and (2) test differences in those factors in males and females.

Methods: The sample was comprised of 411 Portuguese twin pairs of different zygosity aged 12 to 25 years. The SPI and LTPA were assessed with the Baecke questionnaire. Quantitative genetic modeling was used to test alternative models for the presence of additive gene effects (a^2), common or shared environment within the family (c^2), and unique environmental factors (e^2).

Results: The best-fitting models showed sex-specific effects for the two phenotypes. Variance components for SPI in males were a^2=68.4%, c^2=20%, and e^2=11.6%; and in females, a^2=39.8%, c^2=28.4%, and e^2=31.8%. For variation in LTPA, genetic factors in males explained 63%, common environment was not significant, and unique environment explained 37%. In females, contributing factors were a^2=32%, c^2=38%, and e^2=30%.

Conclusions: Genetic effects explained a considerable amount of variation in SPI and LTPA, which were greater in males than in females. The relevance of shared environmental factors (family and peers) and nonshared environmental factors in SPI and LTPA is particularly evident in females.

Medical Subject Headings (MeSH): exercise, genetics, physical fitness, twins (Am J Prev Med 2002;23(2S):87–91)

Introduction

Whatever the measurement approach we choose to assess physical activity, it consistently shows wide variation at the population level. Explanations for this variation have been proposed primarily by physical activity epidemiologists and behavioral science experts. Their major focus has been the identification of correlates to explain this variation, that is, demographic aspects, socioeconomic status, psychological characteristics, and context variables, such as the presence of sports facilities and organizations.[1] Moreover, behavioral experts have grounded their studies in diverse psychosocial theories that often provide less explanatory power than expected. The interplay of the determinants suggested by the models of these theories explain, at best, only 20% to 30% of the total variance in physical activity levels.[2]

From the Faculty of Sport Sciences and Physical Education, University of Porto (Maia), Porto, Portugal; and Faculty of Physical Education and Physiotherapy, Department of Sports and Movement Sciences, Katholieke Universiteit, Leuven (Thomis, Buenen), Leuven, Belgium

Address correspondence and reprint requests to: José A.R. Maia, PhD, Faculty of Sport Sciences and Physical Education, Rua Dr. Plácido Costa, 91, 4200 Porto, Portugal. E-mail: jmaia@fcdef.up.pt.

An often-overlooked approach to elucidate the variation in physical activity level is the genetic epidemiologic framework. This point has been made clear by Bouchard and Rankinen,[3] who note that important publications concerning the interplay of physical activity with morbidity conditions, longevity, and health benefits typically emphasize average effects, such as positive, negative, and effect sizes, with less focus on variability or patterns within this variability. The major interest is, for example, often on mean differences in blood lipids or blood pressure between groups of active and non-active people. These studies do not pay sufficient attention to interindividual differences and group heterogeneity. These latter aspects are the main theme for genetic epidemiologic research—the quantification and interpretation of the genetic and environmental sources of individual differences within a given population.

Family data are used to document phenotype similarities among family members, such as those shown in parent–child correlations and sibling correlations. These correlations, however, reflect a mix of cultural and genetic transmission. On the other hand, studies based on both twin types can provide evidence for

 PII S0749-3797(02)00478-6

genetic transmission. Family data can also be used to look for associations or linkages between a phenotype and a gene or a marker at a specific locus in the genome. No such studies have been yet reported for physical activity levels. Consequently, in this article we concentrate on the results from animal models as well as research on twins.

In mice, daily physical activity[4,5] and circadian rhythms of locomotion[6,7] were investigated, as well as activity behavior in search of food in fruit flies.[8] These studies revealed that genetic factors partly determine the variation in different modalities of physical activity. Furthermore, loci were identified that were linked to physical activity phenotypes. The identification of these loci will likely lead in the future to identifying genes responsible for physical activity levels. No such studies have been reported yet in humans.

Human studies using the twin methodology have focused on physical activity phenotypes assessed by questionnaires. Moderate- and high-intensity scores,[9] general physical activity scores,[10,11] or summed hours of sports participation[12] have been used in these analyses. Despite diversity in sample size, age range, phenotypic physical activity assessment, and statistical procedures in quantifying genetic influences, results indicate in the aggregate that genetic factors are important and account for a substantial portion of the variation at the population level (from 35% to 83%).

In spite of this general trend, data are scarce and have dealt mainly with adult subjects, primarily males, and has used broad or narrow definitions of the physical activity phenotype. Moreover, only one paper used all the possibilities of biometric genetics,[12] which means that the several alternative hypotheses formulations of the interplay of genetic and environmental factors were not typically tested.

Recent reports by Rankinen et al.[13] on genes associated with performance and health-related fitness phenotypes, and Bouchard and Rankinen[3] on heterogeneity in the response to physical activity programs, stress the absence of data on genes that may be responsible for the broad variation encountered at the population level. The major goals of this study are (1) to estimate the relative quantitative contribution of genetic factors, along with shared (i.e., sports or physical activity habits that act in common on twins, siblings, or on parent and child) and nonshared environmental factors (i.e., physical activity or sports habits that are uncorrelated across twins, siblings, or parent and child), to the variation in two phenotypes—sports participation index (SPI), and leisure-time physical activity (LTPA); and (2) and to test for differences in these factors in males and females.

Methods

Data reported in this study are selected from a large population-based study of familial aggregation in physical activity and sports participation habits of the Portuguese population. All subjects were recruited by advertisements in various regions of the country. This study was approved by the Ethics Review Board of the University of Porto, and written informed consent was obtained from the subjects (or their parents). From a sample of 6000 nuclear families, we selected a subsample that included 411 twin pairs aged 12 to 25 years of both sexes. Participants' formal education ranged from secondary to completion of a college degree, and socioeconomic status from middle class to upper-middle class (according to income strata data by the Portuguese Bureau of Statistics). The number of subject pairs in each zygosity and sex group follows: monozygotic males (MZ_m), n=85, age=17.9±6.5; monozygotic females (MZ_f), n=118, age=16.7±5.1; dizygotic males (DZ_m), n=68, age=15.8±5.1; dizygotic females (DZ_f), n=85, age=16.9±5.6; and dizygotic opposite-sex twin pairs (DZ_{os}), n=55, age=17.5±5.9.

The two physical activity phenotypes used in this report are expressed as a quantity with no specific measurement unit (i.e., as an arbitrary unit in an interval scale). The SPI and LTPA were estimated based on the questionnaire by Baecke et al.,[14] which has been shown to be valid and highly reliable.[15,16] The SPI is a composite score of four items that takes into account the expected energy expenditure for a given sport, number of hours practiced per week, and number of months per year. The LTPA is also a composite score based on the following four items: hours of TV watching, frequency of walking in leisure time, number of minutes spent walking per day, and frequency of bicycle riding.

A random sample of 100 twin pairs was used to estimate the reliability of questionnaire responses, especially in the two phenotypes of interest. The intraclass correlation coefficient (R) was used, and the 95% confidence intervals obtained for the SPI and LTPA composite scores were $0.70<R_{SPI}<0.95$; $0.60<R_{LTPA}<0.90$.

Zygosity determination was assessed with the indirect method by Peeters et al.[17] based on telephone interviews with mothers of twins. The method has been validated against DNA fingerprinting and shown to be valid and reliable.[18] The same investigator (JARM) scored all zygosity questionnaires. A random sample of 100 twin pairs was re-rated for zygosity 1 month later. Cohen's kappa was used to test the consistency of classification, and the value was 1, meaning that there were no errors in zygosity classification.

Assessment of the heritability in SPI and LTPA variation was done by using the basic biometric genetic model.[19,20] Total variation (V_{tot}) in these multifactorial traits was comprised of three components ($V_{tot}=V_G + V_C + V_E$): genetic (V_G), common or shared environment within the family (V_C), and unique environmental factors (V_E). In a similar manner, the contribution of environmental factors shared by family members (variance related to common environmental factors, $c^2=V_C/V_{tot}$) and the variance proportion of environmental factors that act on an individual level ($e^2=V_E/V_{tot}$) were estimated. When using this type of additive model, several assumptions are made: no interaction between genes and environment (different genotypes react equally to similar environmental factors); no gene × environment correlation (similar exposure of environments for different genotypes); no gene × gene interaction; and no assortative mating (i.e., mating is random).

Path analysis procedures were used to estimate path coef-

Table 1. Pearson twin-pair correlation coefficients for SPI and LTPA by zygosity and gender type

Zygosity	SPI	LTPA
MZ_m	0.82	0.69
DZ_m	0.46	0.22
MZ_f	0.90	0.72
DZ_f	0.53	0.56
DZ_{os}	0.49	0.31

MZ_m, monozygotic males; DZ_m, dizygotic males; MZ_f, monozygotic females; DZ_f, dizygotic females; DZ_{os}, dizygotic pairs of opposite sex; LTPA, leisure-time physical activity; SPI, sports participation index.

ficients. Correlational and causal paths from and between latent and observed variables were defined and these relationships were expressed in terms of linear equations. Alternative models (e.g., a model including A, C, and E against a model including only C and E to test for evidence of A as a significant source of variation) were tested and their goodness of fit was evaluated by the lowest chi-square ratio test and Akaike's information criterion (AIC), an index that evaluates the parsimony of each model.[20] An iterative procedure estimated the contributions of each causal path and calculated confidence intervals (CIs) for all contributing factors.

Using the model-fitting approach, 16 alternative models were tested for major influences of additive genes (A), dominance, shared environmental influences (C), and nonshared environmental influences (E) on the two phenotypes. The most restrictive model assumed that only unique environmental factors were responsible for all variation. A more complex one assumed that genetic and environmental influences were different by sex.

Sex differences observed in the SPI and LTPA indices can result from differences in the magnitude of genetic influences, environmental influences, or both, or from a different set of genes acting in males and females. Sex heterogeneity was tested in specific models that included the data (in a variance–covariance matrix) of five types of twins (MZ_m, MZ_f, DZ_m, DZ_f, and DZ_{os}).

All statistical analyses were conducted with SPSS, version 10, and Mx.[21]

Results

Exploratory data analysis was conducted on basic statistical assumptions of the twin method. We found no significant birth order effects in means or variances. No departures from normality were found. Pearson correlation coefficients between age and SPI or LTPA in each zygosity and sex group were low and not significant, ranging between 0.10 and 0.15 ($p>0.05$). Correlation coefficients for all phenotypes are shown in Table 1.

According to theoretical expectations,[19,20] the correlation between MZ twins should be 1 and between DZ twins 0.50 (half of the MZ value) if only genetic factors account for the total variation in SPI and LTPA. If correlation coefficients are lower than these values, but still in this proportion (e.g., $r_{MZ}=0.80$ and $r_{DZ}=0.40$), then not only genes but also specific or nonshared environmental factors are also involved. On the other hand, if this proportion no longer exists (e.g., $r_{MZ}=0.80$ and $r_{DZ}=0.60$), then genes, nonshared (unique) environmental factors, and shared environmental factors are all contributing to total variation in these phenotypes.

Table 2. Main results of variance estimates of additive genetic factors ($a^2=V_A/V_{tot}$), shared environment ($c^2=V_C/V_{tot}$), and unique environment ($e^2=V_E/V_{tot}$)

Phenotypes	a^2	c^2	e^2
SPI			
Males (%)	68.4	20.0	11.6
95% CI	(41.5%–89.2%)	(0.0%–46.8%)	(0.8%–15.6%)
Females (%)	39.8	28.4	31.8
95% CI	(0.4%–73.0%)	(0.0%–58.7%)	(23.4%–43.4%)
LTPA			
Males (%)	63.0	0.0	37.0
95% CI	(48.9%–73.3%)	(0.0%–0.8%)	(26.7%–50.2%)
Females (%)	32.0	38.0	30.0
95% CI	(0.4%–61.8%)	(10.0%–61.6%)	(22.7%–40.2%)

CI, confidence interval; LTPA, leisure-time physical activity; SPI, sports participation index.

In the various sex-by-zygosity comparisons, the pattern of correlations was different. For example, for the SPI, $r_{MZm}=0.82$ and $r_{DZm}=0.46$. In contrast, for LTPA, $r_{MZm}=0.69$ and $r_{DZm}=0.22$, which suggests that additive genetic factors, as well as nonshared and shared environmental factors are acting differently on LTPA than on the SPI.

Models with different degrees of complexity were compared for their goodness-of-fit values. Retained models should not only have the lowest chi-square and AIC values, but should also be substantively meaningful in their parameter estimates. A model with specific scalar estimations for genetic (A) and environmental factors (C and E) was found to be the best fitting and most parsimonious of the 16 models (see Table 2). Global fit measures for SPI were $\chi^2=16.27$, $p=0.06$, and AIC$=-1.73$, and for LTPA, $\chi^2=6.23$, $p=0.71$, and AIC$=-11.68$.

In males, genetic factors accounted for 68% of the total SPI variance, whereas in females the estimate was about 40%. A similar trend was observed for LTPA, with genetic factors explaining 63% of the total variance in males and only 32% in females. In males, shared environmental factors were more relevant than nonshared influences for SPI, but in females it was just the opposite. For LTPA, unique environmental influences showed a variance estimate of 37% in males, and shared influences were almost irrelevant. This was not the case in females, as shared environmental variance was 38% and unique influences reached 30%.

Discussion

Individual diversity in any measured behavior should not be regarded as a nuisance or irrelevant noise in any

data set. This is the fulcrum on which systematic differences in the genetic and environmental makeup of individuals can be investigated. This was the approach used in this study to account for the observed variance in SPI and LTPA.

Four factors may have affected the results and the conclusions of the present study. One is the limited age range of our sample—adolescents and young adults. Yet, it should be pointed out that there are no available studies that cover the entire life span. It would be a difficult task to have adequate phenotyping across the entire life span with representative samples of each age group. Available data cover only specific age ranges: adolescence,[12] young adults,[9,22] and adults.[11]The second limitation has to do with the use of questionnaires in assessing physical activity. As is well known in epidemiologic research, the use of more objective devices such as accelerometers or doubly labeled water estimates is almost impossible in large samples. The Baecke questionnaire has been reported as a valid and reliable instrument.[15] More recently, Philippaerts et al.[16] showed its validity against the criterion of doubly labeled water. Third, reliability estimates from respondents to questionnaires are often low. The confidence intervals obtained in this sample are within acceptable limits, and the reported point estimates (R_{SPI}=0.88 and R_{LTPA}=0.82) are very similar to those of Philippaerts et al.[16] (R_{SPI}=0.93 and R_{LTPA}=0.87), as well as to the best estimates available in the literature as reviewed by Montoye et al.[15] and Pereira et al.[23] Fourth, an indirect method of classifying twins as MZ or DZ was used for practical and financial reasons. Classification of twins by zygosity by questionnaire is common practice in genetic epidemiology research. The method was recently shown to be quite concordant compared to a classification based on DNA markers.[18] Misclassification of zygosity is therefore expected to be minimal in this study.

The advantages of quantitative modeling as used in the present study include testing several a priori competitive models that interpret differently the interplay between genetic and environmental factors. With the goals of defining which model fits the data better, quantitative modeling is more parsimonious and is substantively plausible.

Several interesting findings should be highlighted. First is that genetic factors show moderate to high effects in predicting physical activity scores in both SPI and LTPA. These results are consistent with reports in animal studies, although different phenotypes were used. A similar trend was shown in human studies, although adult male twins were used and intraclass correlations and heritabilities were calculated,[9,11] or were based on young subjects and calculated intraclass correlations,[10], or used genetic modeling.[12] This genetic influence on SPI and LTPA could be due to motor and somatic features that are contributing factors in sports and leisure activities.[12,24] It is also known that motor and somatic traits are partly under genetic control (for a review, see Bouchard et al.[24]).

Second, the results show that genetic factors are of differential importance in males and females. Although variance estimates were different in both sexes, this does not imply that different sets of sports and LTPA genes may be acting in males and in females, because the best-fitting model was a specific scalar ACE model (see methodology section). Boomsma et al.[22] found similar results in sports practice in a sample of 90 twin pairs, aged 14 to 20 years, with a_f^2=35% and a_m^2=77%. Beunen and Thomis[12] reported that a specific scalar model was the best fitting their sample of 91 twin pairs aged 15 years, where no evidence was found for a different set of sports activity genes in the two sexes. In contrast, Koopmans et al.,[25] in a large sample of 1582 twins, aged 13 to 22 years, also found no differences in the variance of genetic factors in the sports practices of males and females (a^2=48%, c^2=38%, and e^2=12%).

Environment is an important contribution to the variance of SPI and LTPA. Shared and unique environmental factors had variable estimates not only in SPI and LTPA, but by gender as well. Environmental factors play an important role not only in SPI but also in LTPA. Although social learning and parental role modeling are relevant factors in explaining familial similarity (i.e., shared environmental influences), familial aggregation of SPI and LTPA is a controversial issue in the literature. Some researchers[26,27] found some aggregation in physical activity habits within nuclear families, while others[10,28,29] found little or no familial aggregation. Our results showed that shared estimates are more or less equivalent between sexes for SPI, but divergent for LTPA. It could be suggested that, in males, leisure activities seem to diverge from parental role modeling, and that other influences outside the family may have also an important role to play, such as peers acting differently on each member of the twin pair. In a Portuguese study of family aggregation[30] in physical activity habits of males aged 16 to 19 years, no parental influence was found. In contrast, peers and other social role modelers influenced their physical activity habits. The observed nonshared environmental influences acting differently on each member of a twin pair might suggest a unique or uncorrelated exposure to physical activity and sports. There is a need for a careful description and interpretation of the mechanisms by which environmental factors produce such variation in physical activity levels.

From the results of our study, the following conclusions are possible:

1. There are moderate to high genetic influences on SPI and LTPA.
2. These effects are of different magnitude, being more important in males.

3. Environmental factors are also relevant, especially shared and nonshared environmental influences in females.

Our results broaden the debate about methods available for understanding the influences on physical activity in populations.

Two lines of implications emerge from these results. First, the suggestion of genetic effects in SPI and LTPA implies that not everybody is equally prone to engage in such activities. Consequently, the potential to reduce the risk of some chronic diseases through a physically active lifestyle is not the same for all individuals. Thus, different strategies should be developed for such subgroups in the population, with a major focus on high-risk individuals. Second, the importance of shared and nonshared environmental influences (i.e., family, friends, peers, and other role modelers) suggests that intervention programs could possibly be designed to involve family members and friends.

We are grateful for the thoughtful and helpful comments of Adrian Bauman, Abby King, Karen Calfas, and the anonymous reviewers.

References

1. Sallis J, Owen N. Physical activity and behavioral medicine. Newbury Park, CA: Sage Publications, 1998.
2. Caspersen C, Nixon P, DuRant R. Physical activity epidemiology applied to children and adolescents. Exerc Sport Sci Rev 1998;26:341–403.
3. Bouchard C, Rankinen T. Individual differences in response to regular physical activity. Med Sci Sports Exerc 2001;33(suppl):446–51.
4. DeFries JC, Gervais MC, Thomas EA. Response to 30 generations of selection for open-field activity in laboratory mice. Behav Genet 1978;8:3–13.
5. Flint J, Corley R, DeFries JC, et al. A simple genetic basis for a complex psychological trait in laboratory mice. Science 1995;269:1432–5.
6. Mayeda AR, Hofstetter JR. A QTL for genetic variance in free-running period and level of locomotion activity between inbred strains of mice. Behav Genet 1999;29:171–6.
7. Mayeda AR, Hofstetter JR, Belknap JK, Nurberger JI Jr. Hypothetical quantitative trait loci (QTL) for circadian period of locomotor activity in CXB recombinant inbred strains of mice. Behav Genet 1996;26:505–11.
8. Sokolowski B, Greenspan M, Pereira HS. Natural behavior polymorphism due to a cGMP-dependent protein kinase of drosophila. Science 1997;277:834–6.
9. Lauderdale DS, Fabsitz R, Meyer JM, Sholinsky P, Ramakrishnan V, Goldberg J. Familial determinants of moderate and intense physical activity: a twin study. Med Sci Sports Exerc 1997;29:1062–8.
10. Aarnio M, Winter T, Kujala UM, Kaprio J. Familial aggregation of leisure-time physical activity—a three generation study. Int J Sports Med 1997;18:549–56.
11. Kaprio J, Koskenvuo M, Sarna S. Cigarette smoking, use of alcohol, and leisure-time physical activity among same-sexed adult male twins. In: Gedda L, Parisi P, Nance WE, eds. Twin research Volume 3: proceedings of the Third International Congress on Twin Studies, June 16–20, 1980, Jerusalem. New York: A.R. Liss, 1981:37–46.
12. Beunen G, Thomis M. Genetic determinants of sports participation and daily physical activity. Int J Obes 1999;3:1–9.
13. Rankinen T, Pérusse L, Rauramaa R, Rivera MA, Wolfarth B, Bouchard C. The human gene map for performance and health-related fitness phenotypes. Med Sci Sports Exerc 2001;33:855–67.
14. Baecke JA, Burema J, Fritjers ER. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. Am J Clin Nutr 1982;36:936–42.
15. Montoye HJ, Kemper HCG, Saris WHM, et al. Measuring physical activity and energy expenditure. Champaign, IL: Human Kinetics, 1996.
16. Philippaerts RM, Lefevre J. Reliability and validity of three physical activity questionnaires in Flemish males. Am J Epidemiol 1998;147:982–90.
17. Peeters H, Van Gestel S, Vlietink R, Derom C, Derom R. Validation of a telephone zygosity questionnaire in twins of known zygosity. Behav Genet 1998;28:159–63.
18. Chen WJ, Chang H-W, Wu M-Z, et al. Diagnosis of zygosity by questionnaire and polimarker polymerase chain reaction in young twins. Behav Genet 1999;29:115–23.
19. Falconer DS. Introduction to quantitative genetics, 3rd ed. Essex: Longman Scientific & Technical, 1989.
20. Neale MC, Cardon LR. Methodology for genetic studies of twins and families. Dordrecht, Netherlands: Kluwer Academic Press, 1992.
21. Neale MC. Mx: statistical modeling, 4th ed. Richmond, VA: Department of Psychiatry, Medical College of Virginia, 1997.
22. Boomsma DI, Van der Bree MM, Orlebeke JF. Resemblances of parents and twins in sports participation and heart rate. Behav Genet 1989;19:123–41.
23. Pereira MA, Fitzgerald SJ, Gregg EW, et al. A collection of physical activity questionnaires for health-related research. Med Sci Sports Exerc 1997;29(suppl):S15–18.
24. Bouchard C, Malina R, Pérusse L. Genetics of fitness and physical performance. Champaign, IL: Human Kinetics, 1997.
25. Koopmans JR, van Doornen LPJ, Boomsma DI. Smoking and sports participation. In: Goldbourt U, ed. Genetic factors in coronary heart disease. Dordrecht, Netherlands: Kluwer Academic Press, 1994:217–35.
26. Freedson P, Evenson S. Familial aggregation in physical activity. Res Q Exerc Sport 1991;62:384–9.
27. Moore LL, Lombardi DA, White MJ, Campbell JL, Oliveira SA, Ellison RC. Influence of parents' physical activity levels on activity levels of young children. J Pediatrics 1991;118:215–9.
28. McMurray RG, Bradley CB, Harrell JS, Beruthal PR, Frauman AC, Bangdiwala SI. Parental influences on childhood fitness and activity patterns. Res Q Exerc Sport 1993;64:249–55.
29. Sallis J, Patterson T, Buono M, Atkins C, Nader P. Aggregation of physical activity habits in Mexican-American and Anglo families. J Behav Med 1988;11:31–41.
30. Moreira e Sá MP. Semelhança familiar nos hábitos de actividade física: um estudo em rapazes do 10° ao 12°° ano de escolaridade (Familial resemblance in physical activity habits: a study of boys in the 10th to 12th grades). Master's thesis. University of Porto, Porto, Portugal, 2000.