

# Desafios no estudo de famílias nucleares: etapas iniciais de análise

CDD. 20.ed. 613.7  
614.4

Cláudia Lúcia de Moraes FORJAZ\*  
Teresa BARTHOLOMEU\*  
Januária Andréa Souza REZENDE\*\*  
Jorge Alberto de OLIVEIRA\*  
Luciano BASSO\*  
Go TANI\*  
Antonio PRISTA\*\*\*  
José António Ribeiro MAIA\*\*\*\*

\*Escola de Educação Física e Esporte, Universidade de São Paulo.

\*\*Escola Superior de Educação Física de Muzambinho.

\*\*\*Faculdade de Ciências de Educação e Desporto, Universidade Pedagógica de Moçambique - Moçambique.

\*\*\*\*Faculdade de Desporto, Universidade do Porto - Portugal.

## Resumo

Os estudos em famílias nucleares possibilitam a avaliação da existência de agregação familiar numa dada característica, permitindo avaliar o quanto da variação dessa característica na população pode ser atribuída à variação genética existente entre os sujeitos. Os resultados desses estudos possibilitam a elaboração de estratégias de intervenção mais eficientes e direcionadas, além de serem o ponto de partida para estudos mais complexos de epidemiologia genética, como a identificação de genes responsáveis pela característica em análise. Nas áreas de Ciências do Desporto e Educação Física, alguns estudos têm sido realizados para verificar a existência de agregação familiar e a influência genética em traços relacionados a estados de saúde, atividade e desempenho físicos. Entretanto, esses esforços revelam-se escassos em populações lusófonas e são, praticamente inexistentes, na população brasileira. Dessa forma, o presente artigo teve como propósito abordar, de forma simples e introdutória, aspectos importantes de estudos em famílias nucleares. Para tanto, foram analisadas as duas fases iniciais dos estudos de Epidemiologia Genética, ou seja, a identificação e quantificação da agregação familiar e da heritabilidade. As diferentes etapas de análise foram exemplificadas com dados reais, coletados em famílias nucleares pertencentes a um estudo realizado na cidade de Muzambinho-MG. Espera-se que esse artigo forneça subsídios e estímulo aos pesquisadores iniciantes neste tipo de investigação.

UNITERMOS: Agregação familiar; Heritabilidade; Perímetro da cintura; Volume semanal de atividade física; Muzambinho.

## Introdução

Numa pesquisa científica é comum partir-se de problemas complexos para descrever as variáveis em estudo e interpretar os seus significados. Essas variáveis, geralmente contínuas, possuem um atributo essencial - têm valores distintos entre sujeitos, apresentando uma grande variabilidade de expressão numa dada população (BOUCHARD, MALINA & PÉRUSSE, 1997; MAIA, SEABRA,

GARGANTA, CHAVES, SOUZA & SANTOS, 2010a). Por exemplo, em 204 adultos de ambos os sexos da cidade de Muzambinho, o perímetro da cintura (importante indicador de risco cardiovascular) variou de 66 a 125 cm, enquanto o volume semanal de atividade física apresentou resultados que se situavam entre 0 e 5400 min/semana<sup>1</sup> (FIGURA 1).

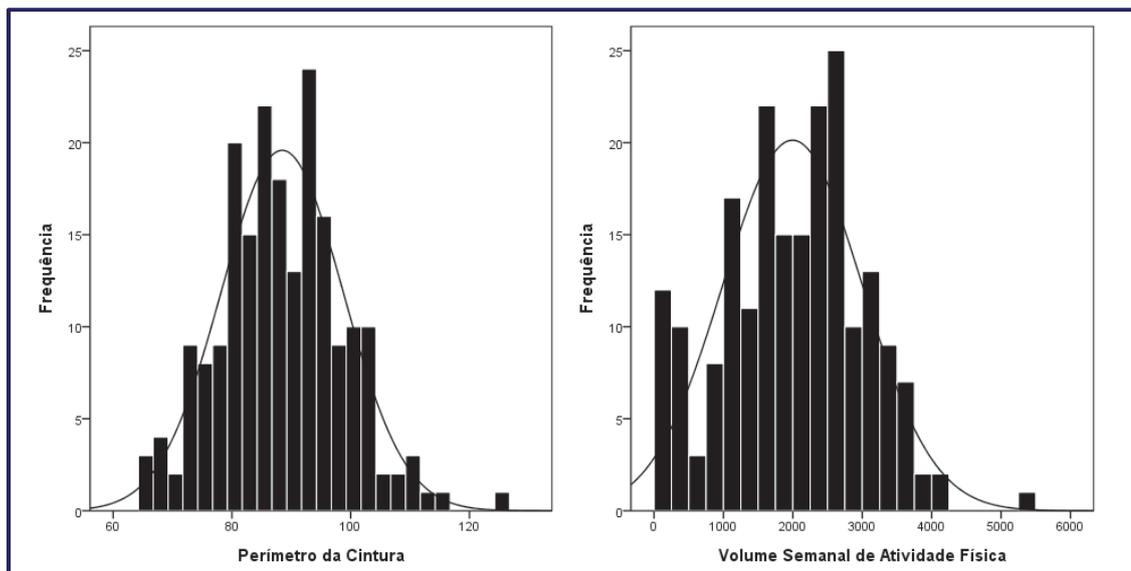


FIGURA 1 - Variabilidade inter-individual na distribuição dos valores do perímetro da cintura (cm) e volume de atividade física (min/semana) em adultos da cidade de Muzambinho.

A variação quantitativa ocorrida num determinado traço (fenótipo ou característica) parece depender de um conjunto de fatores denominados, genericamente, de genéticos e ambientais, sendo relevante quantificar a participação relativa de cada um na explicação da variância total da característica em análise (BEUNEN & THOMIS, 1999; BOUCHARD, MALINA & PÉRUSSE, 1997; BURTON, TOBIN & HOPPER, 2005; FEITOSA & KRIEGER, 2002; MAIA et al., 2010a; RAO, 2001; RICE & BORECKI, 2001). Se a variância de um dado fenótipo for explicada na sua totalidade (i.e.100%) por fatores genéticos (como são, por exemplo, as características de algumas doenças monogênicas raras), intervenções ambientais (como, por exemplo, a dieta ou os fármacos) têm pouca chance de conseguir modificá-lo. Porém, na medida em que o componente ambiental ganha maior contribuição relativa na expressão da característica, as manipulações do ambiente passam a ter maior relevância. O conhecimento preciso e esclarecido acerca dos modos de transmissão de determinadas características permite a identificação de subconjuntos de sujeitos aparentados com maior ou menor propensão a apresentar aquela característica, possibilitando a intervenção atempada em indivíduos de maior risco.

A Epidemiologia Genética (EG) é uma área de estudo confluyente entre a Epidemiologia, a Genética Humana e a Estatística, que se ocupa, genericamente, dos determinantes de um conjunto multifacetado de fenótipos descritores de estados patológicos (i.e., doenças) e de saúde, fazendo a ligação entre fatores genéticos e ambientais

transmitidos ao longo de gerações (BURTON, TOBIN & HOPPER, 2005; FEITOSA & KRIEGER, 2002; MAIA et al., 2010a). A exploração interpretativa oriunda da EG percorre sequencialmente diferentes etapas (BEUNEN & THOMIS, 1999; BOUCHARD, MALINA & PÉRUSSE, 1997; BURTON, TOBIN & HOPPER, 2005; FEITOSA & KRIEGER, 2002; MAIA et al., 2010a; RICE & BORECKI, 2001):

1) Estudos de agregação familiar - analisam a presença de associações significantes (razão de chances, correlações intraclasse, análise de variância) de um determinado fenótipo entre membros de uma família com diferentes graus de parentesco;

2) Estudos de heritabilidade - quantificam a magnitude da contribuição relativa da agregação familiar expressa pelos fatores genéticos num dado fenótipo;

3) Estudos de "linkage" - pesquisam a cossegregação (cotransmissão) de marcadores genéticos em sujeitos aparentados com o propósito de identificar regiões cromossômicas que contêm genes candidatos para explicar a associação evidenciada;

4) Estudos de associação - avaliam a magnitude da relação entre um determinado gene candidato e respectivos alelos e a ocorrência do fenótipo em causa em indivíduos geralmente não aparentados, num delineamento clássico de caso-controle.

Este artigo aborda as duas primeiras etapas, que são passos essenciais num estudo de EG, pois a investigação acerca da arquitetura genética de um fenótipo só se justifica perante evidências inequívocas da existência de agregação familiar e de contribuição genética substancial nessa agregação (RAO, 2001).

Na literatura internacional, alguns pesquisadores têm avaliado as contribuições relativas - genéticas e ambientais - em diversos fenótipos de saúde e atividade física. A heritabilidade ( $h^2$ ) - proporção da variância total que pode ser atribuída a fatores genéticos - de diferentes indicadores de gordura corporal, por exemplo, foi reportada em 19 estudos (FERMINO, GARAGANTA, SEABRA & MAIA, 2007), enquanto nove trabalhos estimaram o contributo genético para os níveis de atividade física (CHAVES, SOUZA, SANTOS, GARGANTA, SEABRA & MAIA, 2010).

A pesquisa centrada no paradigma da EG no território das Ciências do Desporto e Educação Física é relativamente reduzida no espaço lusófono, sendo praticamente inexistente no Brasil. Na base de dados MedLine, só foi possível localizar sete trabalhos relacionados com distintos marcadores de saúde cardiovascular (COLLETO, CARDON & FULKER, 1993; DE OLIVEIRA, PEREIRA, DE ANDRADE, SOLER & KRIEGER, 2008; KRIEGER, MORTON, RAO & AZEVEDO, 1980; PROVINCE & RAO, 1985; RAO, MACLEAN, MORTON & YEE, 1975; ROBINSON, BORGES-OSORIO, CALLEGARI-JACQUES, ACHUTTI, DA SILVEIRA, KLEIN & COSTA, 1991; VELASQUEZ-MELENDZ, PARRA, GAZZINELLI, WILLIAMS-BLANGERO & CORREA-OLIVEIRA, 2007) e nenhum sobre aptidão física ou atividade física, conduzido em uma população brasileira (pesquisa realizada em junho de 2010).

Dessa forma, o propósito do presente artigo foi abordar, de forma simples e introdutória, aspectos

importantes de estudos de agregação familiar e heritabilidade, visando fornecer subsídios e estímulo aos iniciantes neste tipo de investigação.

### **Estrutura do artigo**

Para facilitar a compreensão, o artigo foi dividido em cinco partes, que percorrem os aspectos relacionados: 1) ao delineamento de um estudo e coleta dos dados; 2) à tabulação da informação; 3) à análise exploratória inicial e normalização das distribuições; 4) à representação gráfica da agregação familiar e suas medidas descritivas; e 5) à quantificação e interpretação da heritabilidade.

Partiu-se de dados providos de 138 famílias nucleares da cidade de Muzambinho-MG, que participaram de um estudo temático cujos objetivos e características foram anteriormente descritos (BASSO, JUNIOR, OLIVEIRA, FORJAZ, SOUZA, PRISTA, MAIA & TANI, 2009). Esse estudo teve a aprovação do Comitê de Ética da Escola de Educação Física e Esporte da USP (processo CEP-EEFE-USP: CAAE-0033.0.342.000-07). Os dados foram coletados de abril/08 a novembro/09.

Dois variáveis foram abordadas: o perímetro da cintura (C), que foi medido na altura da cicatriz umbilical, e o volume semanal de atividade física ( $V_{AF}$ ), que foi avaliado a partir da informação coletada numa entrevista sobre práticas físicas realizadas em diversos contextos: locomoção, trabalho, escola, recreio e lazer.

## **Delineamento de estudos e coleta de dados**

Num delineamento centrado na agregação familiar e heritabilidade, a viabilidade da execução, o poder estatístico e a relação custo-efetividade devem ser considerados (RAO, 2001). Para tanto, aspectos relativos à amostragem, fenotipagem e procedimentos analíticos precisam ser cuidadosamente planejados (FEITOSA & KRIEGER, 2002).

Os estudos de agregação familiar e heritabilidade englobam, geralmente, famílias nucleares (pais e filhos), "pedigrees" extensos (filhos, pais, avôs, tios, primos, etc.) ou gêmeos mono e dizigóticos (BOUCHARD, MALINA & PÉRUSSE, 1997; RICE & BORECKI, 2001). O delineamento envolvendo famílias nucleares, i.e. duas gerações, permite avaliar aspectos da agregação familiar de uma determinada característica, mas não fraciona, da variância total, a parte que se refere aos fatores genéticos daquela relativa ao ambiente cultural da família. A diferenciação dessas contribuições pode

ser obtida em estudos de pedigrees extensos a partir da inclusão de um novo parâmetro no modelo - um índice de ambiente compartilhado - elaborado com base na obtenção de várias informações provenientes da família, e que é designado por "household effect" (BOUCHARD, MALINA & PÉRUSSE, 1997; BURTON, TOBIN & HOPPER, 2005; RICE & BORECKI, 2001). Entretanto, a amostragem em "pedigrees" extensos é de difícil obtenção, havendo normalmente perda substancial de informação (dados omissos), o que torna difícil estimar a magnitude e significado do "household effect" (BOUCHARD, MALINA & PÉRUSSE, 1997; RICE & BORECKI, 2001). Torna-se, portanto, importante conhecer as limitações de cada procedimento de amostragem a fim de se utilizar ferramentas estatísticas para minimizar os problemas associados à estimação de parâmetros e respectiva precisão, bem como a sua consideração na interpretação dos resultados.

O número de sujeitos necessários deve ser definido em razão do poder estatístico do estudo, que leva em consideração a precisão da medição do fenótipo; deve-se considerar também a viabilidade, os custos e a efetividade (FEITOSA & KRIEGER, 2002; RAO, 2001). Em estudos de famílias nucleares, nem sempre é fácil se obter amostras de grandes dimensões, i.e., centenas de famílias. Assim, algumas vezes as famílias são recrutadas em virtude de terem membros portadores dos fenótipos desejados o que, se por um lado aumenta o poder estatístico, por outro viola, num certo sentido, a noção de aleatoriedade da seleção. Contudo, quando se lida com famílias, esse pressuposto pode ser relaxado em função dos propósitos da pesquisa, sem a perda de seu poder de generalização. Há métodos de correção para amostras selecionadas, designados por correção para “ascertainment” (BURTON, TOBIN & HOPPER, 2005).

No estudo das famílias de Muzambinho, as crianças-alvo (que não foram escolhidas em função de nenhuma particularidade) foram as que tinham 10 anos entre os alunos de escolas públicas sorteados para o projeto temático anteriormente descrito. Foram estudadas 251 (135 meninos e 116 meninas) crianças (4,4% dos habitantes de cinco e 18 anos da cidade) (BRASIL. MEC. INEP, 2009), que pertenciam a 138 famílias, das quais foram avaliados 95 pais e 129 mães.

O fenótipo a ser investigado precisa ser definido e medido com grande precisão. Um problema frequente no estudo de famílias se relaciona à métrica das variáveis face ao efeito geracional em causa. Por exemplo, a aptidão aeróbia de pais e filhos pode ser medida pelo teste de uma milha nas crianças e o teste de Cooper nos adultos, o que gera variáveis que expressam, indiretamente, o mesmo aspecto

fisiológico mas com unidades diferentes (tempo e distância, respectivamente). A comparação entre gerações, portanto, torna necessária a equiparação dessa variável, o que dificulta, de algum modo, a análise dos dados. Procedimentos de avaliação da qualidade dos dados, como a execução de medidas múltiplas e a obtenção de informações suplementares de reprodutibilidade da medida são altamente desejáveis (FEITOSA & KRIEGER, 2002).

A identificação de possíveis covariáveis (características que podem influenciar a variabilidade do fenótipo em questão) bem como a determinação da quantidade de variância que extraem do fenótipo em estudo são necessárias. Essa etapa permite obter um refinamento adequado do fenótipo (BOUCHARD, MALINA & PÉRUSSE, 1997). Nos estudos de agregação familiar e heritabilidade, a idade, o sexo e suas interações são covariáveis importantes. É também provável que outras características dos sujeitos possam ser coinfluentes. Por exemplo, o efeito do índice de massa corporal num estudo sobre o perímetro da cintura deve ser considerado.

Uma vez definido o delineamento do estudo, estrutura-se e executa-se o plano de trabalho: obtenção de parcerias e permissões; esclarecimento aos envolvidos; treinamento da equipe; e obtenção de consentimento dos sujeitos.

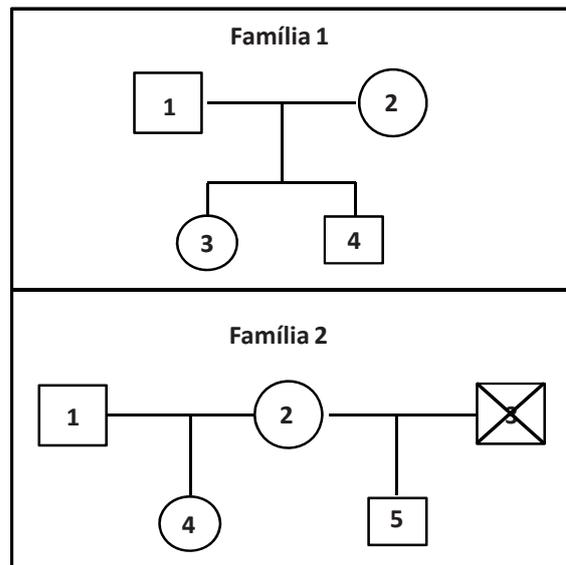
A qualidade dos resultados depende, sobretudo, da coleta dos dados. O local deve permitir a realização adequada de cada medida, a preservação do anonimato da família e o conforto da mesma. Os instrumentos de anotação precisam ser de fácil compreensão e manuseio. Para minimizar as desistências, comuns neste tipo de estudo, estratégias de motivação e busca de perdas devem ser instituídas.

## Tabulação de dados

No estudo completo das famílias de Muzambinho foram avaliadas cerca de 40 variáveis nos 475 sujeitos, totalizando mais de 19000 células de dados. Para minimizar a chance de erros de entrada, a dupla digitação em “softwares” adequados, como o Access (Office for Windows), e a comparação dos dois arquivos em programas como o EpiInfo (UNITED STATES

DEPARTMENT OF HEALTH AND HUMAN SERVICES, 2005) são indicadas.

A estrutura familiar, conhecida como árvore familiar ou “pedigree”, deve ser estabelecida e se compõe da descrição de quem são os pais (fundadores), os filhos e a relação de parentesco biológico entre eles. Para exemplificar, tomemos por base as famílias apresentadas na FIGURA 2.



Quadrado = masculino;  
círculo = feminino;  
X = membro ausente  
(sem dados).

FIGURA 2 - Esquema da estrutura de duas famílias nucleares avaliadas no estudo de Muzambinho.

Uma forma de tabulação desses pedigrees, chamada de formato “linkage”, é a colocação dos dados de cada membro da família em linhas de tabulação e a utilização de cinco colunas para definir a estrutura familiar: família, posição, pai, mãe e sexo. Esse formato é utilizado em praticamente todos os “softwares” de EG. Como se observa na TABELA 1, na coluna família, estabelece-se a que núcleo familiar pertence cada sujeito. Na coluna posição, faz-se para cada família uma numeração contínua, seguindo das gerações ascendentes para as descendentes e iniciando-se com o ascendente do sexo masculino.

Num estudo de famílias, todos os descendentes precisam ter relações familiares estabelecidas, mesmo que não haja dados de um dos fundadores, sendo necessário criar uma linha de tabulação para o pai ou a mãe ausente, como na família 2, o pai 3. As colunas pais e mães estabelecem o vínculo biológico (quem é pai de quem) a partir da identificação do número de posição dos pais e mães de cada sujeito. No caso de famílias nucleares, os fundadores não têm ascendentes e, portanto, coloca-se zero ou deixa-se em branco essas colunas. Na coluna sexo, no exemplo, 1 é homem e 2 mulher.

TABELA 1 - Planilha do Access com a tabulação do “pedigree” das duas famílias da FIGURA 2.

Família	Posição	Pai	Mãe	Sexo	C	V <sub>AF</sub>
1	1	0	0	1	77	2300
1	2	0	0	2	78	1600
1	3	1	2	2	52	1600
1	4	1	2	1	60	1200
2	1	0	0	1	89	1700
2	2	0	0	2	100	600
2	3	0	0	1		
2	4	1	2	2	73	1400
2	5	3	2	1	77	600

C = valor do perímetro da cintura (cm);  
V<sub>AF</sub> = volume semanal de atividade física (min/semana).

As demais colunas da tabulação devem conter os fenótipos (contínuos ou binários) de interesse, bem como todas as covariáveis. No exemplo, são apresentadas variáveis numéricas contínuas. Os dados de C, por exemplo, poderiam ser classificados em normais (1) ou alterados (2) numa nova coluna. No caso de

variáveis com mais de duas categorias, essas precisam ser transformadas em binárias, visto que não estão disponíveis “softwares” que estimem contributos genéticos de variáveis com mais de duas categorias, i.e., variáveis numa escala ordinal. Quando há dados omissos, a célula deve ficar em branco, sem dados.

## Análise exploratória de dados

A análise exploratória dos dados pode ser feita em qualquer “software” de análise estatística (SPSS, SAS, STATA) e tem por finalidade: identificar valores anormais, verificar a presença de valores extremos, analisar a dimensão e estrutura da informação omissa, e verificar a normalidade da distribuição de cada variável. Valores máximos e mínimos incoerentes com valores biológicos devem ser revistos. A presença de valores extremos (“outliers”) deve ser, preferencialmente, inspecionada nos diagramas de caule e folhas (do inglês “stem and leaf”) e no gráfico de extremos e quartis (do inglês “box-plot”). Os extremos identificados devem ser inicialmente avaliados quanto à sua exatidão e, posteriormente, deve-se decidir por sua exclusão, substituição ou permanência. Dados omissos em excesso reduzem o poder dos testes estatísticos e a generalização dos resultados, enquanto a distribuição não aleatória dessas omissões fere pressupostos importantes da análise, sendo necessário identificar sua causa e considerá-la nos cálculos e na interpretação dos resultados. Há algum esforço de estudo (FRIDLEY & DE ANDRADE, 2008) sobre a influência de métodos

de imputação de dados no viés das estimativas obtidas em estudos de famílias, mas a sua discussão foge ao escopo deste artigo.

A normalidade da distribuição dos dados é uma dos pressupostos para a análise da agregação familiar e da heritabilidade. Ela pode ser identificada graficamente (histogramas com curva normal da FIGURA 1) ou por testes formais, como por exemplo o de Kolmogorov-Smirnov. Sempre que há violação desse pressuposto, deve-se recorrer a transformações adequadas em qualquer programa de estatística. Existem “softwares”, como o STATA (STATA CORP, 1996), que têm automatizado diferentes tipos de transformações (logarítmica, inversa, entre outras) e testam, formalmente, a mais adequada em cada situação. Além disso, alguns “softwares” de EG possuem procedimentos internos para resolver o problema da normalização da distribuição dos dados. Por exemplo, o “software” SOLAR (SOUTHWEST FOUNDATION FOR BIOMEDICAL RESEARCH - SFBR, 1995) tem implementado um método baseado no inverso da distribuição normal para normalizar as variáveis (ALMASY & BLANGERO, 1998).

## Análise de agregação familiar

A agregação familiar numa dada característica acontece quando a semelhança dessa característica entre os seus membros é maior do que a esperada por mero acaso (BEUNEN & THOMIS, 1999; BOUCHARD, MALINA & PÉRUSSE, 1997; BURTON, TOBIN & HOPPER, 2005; FEITOSA & KRIEGER, 2002; MAIA et al., 2010a). Essa agregação reflete, na sua essência, a similaridade genética e cultural (por exemplo, hábitos de vida, crenças, conceitos) dentro do seio de cada família (BEUNEN & THOMIS, 1999; BOUCHARD, MALINA & PÉRUSSE, 1997; BURTON, TOBIN & HOPPER, 2005; FEITOSA & KRIEGER, 2002; MAIA et al., 2010a).

Uma etapa prévia à identificação da agregação familiar reclama uma inspeção detalhada à estrutura das famílias, o que pode ser realizada em diferentes “softwares”, dos quais o PEDSTATS (WIGGINTON & ABECASIS, 2005) é um bom exemplo. Como se observa na FIGURA 3 (“output” do PEDSTATS), as 138 famílias de Muzambinho são formadas por 537 pessoas, sendo 286 fundadores e 254 mulheres. Em média, apresentam 3,89 membros, variando de três a nove. Como se trata de um estudo de famílias nucleares há apenas duas gerações presentes. O “output” contém também estatísticas básicas dos fenótipos  $C$  e  $V_{AP}$ , além das possíveis covariáveis, no caso, sexo, idade e índice de massa corporal (IMC).

```

PEDIGREE STRUCTURE
=====
Individuals: 537
  Founders: 286 founders, 251 nonfounders
  Gender: 254 females, 283 males
  Families: 138

Family Sizes
  Average: 3.89 (3 to 9)
  Distribution: 3 (41.3%), 4 (41.3%) and 5 (11.6%)

Generations
  Average: 2.00 (2 to 2)
  Distribution: 2 (100.0%), 0 (0.0%) and 1 (0.0%)

Females Per Family
  Average: 1.84 (1 to 6)
  Distribution: 1 (40.6%), 2 (40.6%) and 3 (15.2%)

Males Per Family
  Average: 2.05 (1 to 6)
  Distribution: 2 (44.9%), 1 (28.3%) and 3 (23.2%)

Checking family connectedness ...
  All individuals in each family are connected.

QUANTITATIVE TRAIT STATISTICS
=====
ALL DATA:
  [All Phenotypes]
  IDADE 475 88.5% 6.420 65.170 23.540 191.332 -0.008
  IMC   438 81.6% 12.700 45.700 21.786 27.882 0.151
  C     438 81.6% 48.000 125.000 76.387 233.068 0.149
  VAF   434 80.8% 0.000 5400.000 1556.820 1042971 0.341
  Total 1785 83.1%

  [Founders Only]
  IDADE 224 78.3% 24.500 65.170 37.120 44.338 -
  IMC   202 70.6% 16.839 45.700 25.689 16.954 -
  C     202 70.6% 65.500 125.000 88.460 105.780 -
  VAF   204 71.3% 0.000 5400.000 1994.890 1020944 -
  Total 832 72.7%

If you find this program useful in your work, please cite:
Wigginton, JE and Abecasis, GR (2005) PEDSTATS: descriptive
statistics, graphics and quality assessment for gene mapping data.
Bioinformatics. 21(16):3445-3447

```

FIGURA 3 - Descrição da estrutura e dos dados das famílias avaliadas no estudo de Muzambinho. Análise pelo PEDSTATS.

Um procedimento gráfico adequado para mostrar a presença de agregação é o referido na FIGURA 4, que foi construído no “software” HLM 6.02 (RAUDENBUSH, BRYK & CONGDON, 2004). Os valores da variável nos diferentes elementos de cada família, já ajustados para as covariáveis significantes, são representados por uma caixa de extremos e quartis. As caixas são

ordenadas em função dos valores medianos de cada família, na esperança de se encontrar um comportamento ascendente, como o observado nos dois painéis da FIGURA 4, o que evidencia a presença de maior variação entre as famílias do que dentro das famílias e, portanto, sugere a agregação familiar relevante nos fenótipos estudados (MAIA et al., 2010a).

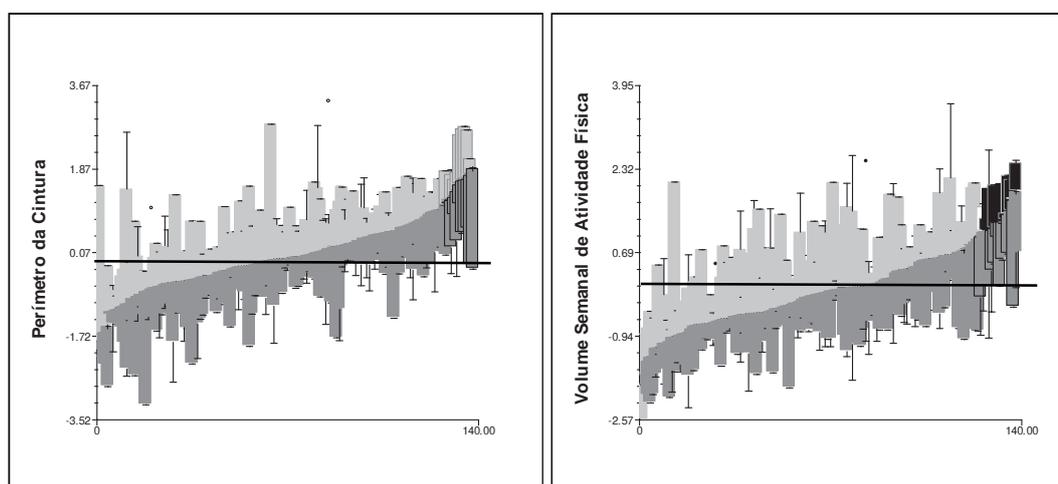


FIGURA 4 - Diagramas de extremos e quartis ordenados pelo valor da mediana dos resultados de perímetro da cintura (cm) e volume semanal de atividade física (min/semana) obtidos em 138 famílias de Muzambinho e previamente ajustados para covariáveis estatisticamente significantes. Gráficos do HML 6.02.

A análise quantitativa da agregação familiar avalia o quanto da variância total ( $V_T$ ) do fenótipo se deve à variância entre famílias ( $V_F$ ). Quando há agregação, a  $V_F$  deve ser maior que variação residual ou intrafamílias, o que pode ser testado por um modelo de ANOVA com efeitos aleatórios ou pelo cálculo da correlação

inter-cluster (BEUNEN & THOMIS, 1999; BURTON, TOBIN & HOPPER, 2005). Na TABELA 2, o valor da razão de F das duas variáveis é significativa e maior do que um, o que reflete a existência de agregação familiar nos fenótipos em causa, i.e. há cerca de duas vezes mais variação entre famílias do que intrafamília.

TABELA 2 - Resultados da ANOVA, comparando as famílias do estudo de Muzambinho em relação aos valores ajustados do perímetro da cintura (C, cm) e volume total de atividade física (VAF, min/semana).

C = ajustamento prévio para idade, sexo, idade<sup>2</sup>, sexo\*idade e índice de massa corporal;  
V<sub>AF</sub> = ajustamento prévio para idade e sexo\*idade.

		Soma dos Quadrados	gl	Quadrado médio	F	Sig
C	Entre famílias	195,09	137	1,424	1,803	0,000
	Intra famílias	236,91	300	0,790		
	Total	432,00	437			
V <sub>AF</sub>	Entre famílias	191,57	137	1,429	1,729	0,000
	Intra famílias	239,43	296	0,809		
	Total	431,00	433			

Essa agregação também é evidenciada quando os coeficientes de correlação entre os membros da família são maiores que zero (BEUNEN & THOMIS, 1999; BOUCHARD, MALINA & PÉRUSSE, 1997; BURTON, TOBIN & HOPPER, 2005; MAIA et al., 2010a; RICE & BORECKI, 2001). Esses coeficientes variam entre os diferentes pares de parentes em virtude da relação genética existente entre eles. Assim, espera-se que a correlação seja elevada entre gêmeos monozigóticos. Em pares de parentes que partilham, em média, metade dos genes idênticos por descendência (pais-filhos, gêmeos dizigóticos e irmãos), essa correlação deve ser metade da observada

nos monozigotos. Correlações maiores que as esperadas indicam a influência do ambiente familiar comum, o que também é sugerido pela existência de correlação significativa entre mãe e pai, que não compartilham informação genética (BEUNEN & THOMIS, 1999).

O cálculo das correlações pode ser feito por diversos “softwares”, como por exemplo, o SAGE (CASE WESTERN RESERVE UNIVERSITY, 2009). O teste formal dos resultados é feito pela divisão do valor obtido nas correlações pelo seu erro padrão, sendo a significância aceita quando o resultado for superior a 1,96.

Um exemplo pode ser observado na TABELA 3. Considerando-se a variável C, as correlações variaram de 0,105 a 0,390, sendo significantes entre irmã-irmão, irmã-irmã e pai-mãe. A ausência de significância nos diferentes pares formados por fundadores e descendentes pode ser parcialmente explicada pelo tamanho

amostral, visto que se a correlação fosse calculada para os dados desses pares em conjunto, a dimensão amostral seria bem maior e o coeficiente de correlação ( $r = 0,157$ ) passaria a ser significativa. Esse resultado realça o aspecto anteriormente discutido sobre a importância do tamanho da amostra (KEEN & ELSTON, 2003).

TABELA 3 - Correlações entre diferentes pares de membros das famílias de Muzambinho em relação ao perímetro da cintura (C) e o volume semanal de atividade física ( $V_{AF}$ ). Os dados de C foram ajustados previamente às variáveis idade, idade<sup>2</sup>, sexo\*idade<sup>2</sup> e índice de massa corporal. Os dados de  $V_{AF}$  foram previamente ajustados para idade e idade<sup>2</sup>.

	C	$V_{AF}$
Pai-Mãe	0,263 ± 0,116 *	0,392 ± 0,104 *
Pai-filho	0,105 ± 0,121 ns	0,055 ± 0,127 ns
Mãe-filho	0,127 ± 0,098 ns	0,065 ± 0,107 ns
Pai-filha	0,242 ± 0,131 ns	0,192 ± 0,130 ns
Mãe-filha	0,196 ± 0,107 ns	0,201 ± 0,101 *
Irmão-irmão	0,231 ± 0,182 ns	0,430 ± 0,159 *
Irmã-irmão	0,320 ± 0,142 *	0,484 ± 0,117 *
Irmã- Irmã	0,390 ± 0,178 *	0,093 ± 0,196 ns

Dados = Média ± erro padrão;  
\* p < 0,05.

Avaliando-se os resultados relativos ao  $V_{AF}$ , observa-se que as correlações entre os diferentes pares variaram de 0,055 a 0,484 e foram significantes nos pares: mãe-filha, irmão-irmão, irmã-irmão e pai-mãe. Além do tamanho amostral, outro aspecto que pode ser levantado para a discussão da grande variedade de resultados entre as correlações de parentes de primeiro grau, diz respeito à precisão do indicador utilizado para se avaliar o fenótipo. Qualquer técnica de medição tem limitações. No exemplo, o  $V_{AF}$  foi avaliado por entrevista, o que implica em imprecisões decorrentes da capacidade

de lembrança do entrevistado, do entendimento das questões formuladas, da clareza da resposta e da interpretação feita pelo avaliador. Embora cuidados sejam tomados para padronizar esses aspectos, na maioria das vezes as limitações são irremediáveis. De fato, elas podem reduzir, em maior ou menor grau, a precisão dos fenótipos, afetando a magnitude e significância das correlações obtidas.

É interessante observar ainda no exemplo, que a correlação entre mãe e pai foi significativa nas duas variáveis, o que sugere a influência do ambiente compartilhado pelos progenitores.

## Cálculo da heritabilidade

Quando uma característica se agrega em famílias, o próximo passo é estimar, quantitativamente, a contribuição genética, ou seja, a heritabilidade ( $h^2$ ) desse fenótipo (BEUNEN & THOMIS, 1999; BOUCHARD, MALINA & PÉRUSSE, 1997; BURTON, TOBIN & HOPPER, 2005; MAIA et al., 2010a; RICE & BORECKI, 2001).

O cálculo da  $h^2$  se baseia num modelo teórico explicativo dos fatores de influência na variabilidade de fenótipos complexos. Num modelo simplificado, a  $V_T$  compreende, basicamente, dois componentes: a variância genética ( $V_G$ ) e a ambiental ( $V_A$ ). Num segundo momento, a  $V_A$  pode ser segmentada em

ambiente compartilhado pela família ( $V_C$ ) e ambiente único ou específico de cada sujeito ( $V_E$ ). Assim:  $V_T = V_G + V_C + V_E$  (BEUNEN & THOMIS, 1999; BOUCHARD, MALINA & PÉRUSSE, 1997; BURTON, TOBIN & HOPPER, 2005; MAIA, SEABRA, SILVA, FREITAS, GOUVEIA & BASSO, 2010b; RICE & BORECKI, 2001). As contribuições relativas de cada componente são estimadas pela razão entre suas variâncias específicas e a variância total. Por exemplo:  $h^2 = V_G/V_T$ .

O cálculo da  $h^2$  assume alguns pressupostos: 1) efeito genético aditivo (múltiplos genes somam pequenos efeitos a  $V_T$ ); 2) inexistência de interação entre ambiente e



A mesma análise é apresentada na FIGURA 6 para a variável C. No entanto, essa análise requer mais cuidado. No modelo 1, considerando-se as covariáveis “clássicas” (idade, sexo e suas derivadas), identifica-se claramente um problema de curtose e a  $h^2$  não é significativa. Por outro lado, incluindo-se o IMC como covariável, a curtose torna-se normal e a estimativa de  $h^2$  aumenta consideravelmente (11 para 31%), tornando-se significativa. Ressalta-se,

no entanto, que a inclusão do IMC fez com que 92% da  $V_T$  fosse explicada pelas covariáveis, o que reduziu a magnitude de variância do fenótipo em causa. Por ser uma medida relativa, a magnitude da  $h^2$  varia consideravelmente em função da magnitude de  $V_T$  na população estudada e das covariadas que são incluídas na análise e, por esses motivos, esses aspectos precisam ser vistos com cautela (BURTON, TOBIN & HOPPER, 2005).

```

*****
*                               Summary of Results - Model 1                               *
*****
H2r is 0.1087918  p = 0.1298231  (Not Significant)
H2r Std. Error: 0.0996230

      Age  p = 4.9717371e-57  (Significant)
      Sex  p = 0.6981629    (Not Significant)
      Age^2 p = 2.7612713e-12  (Significant)
      Age*Sex p = 0.1389135  (Not Significant)
      Sex*Age^2 p = 0.3050713  (Not Significant)

The following covariates were removed from final models: Sex Age*Sex Sex*Age^2

Proportion of Variance Due to All Final Covariates Is  0.6049129

Warning! Residual Kurtosis is 1.2476 which is too high.
*****

*****
*                               Summary of Results - Model 2                               *
*****
H2r is 0.3134728  p = 0.0003794  (Significant)
H2r Std. Error: 0.0961624

      Age  p = 1.4411139e-22  (Significant)
      Sex  p = 0.0105633    (Significant)
      Age*Sex p = 0.0077731  (Significant)
      Sex*Age^2 p = 0.5813321  (Not Significant)
      Age^2 p = 5.1288684e-06  (Significant)
      BMI  p = 2.4589285e-157  (Significant)

The following covariates were removed from final models:  Sex*Age^2

Proportion of Variance Due to All Final Covariates Is:  0.9226688

Residual Kurtosis is 0.5894, within normal range
*****

```

FIGURA 6 - Parte do “output” do SOLAR 4.01 para a variável perímetro da cintura (C). Modelo 1 - Considerando-se as covariáveis: sexo, idade, idade<sup>2</sup>, sexo\*idade e sexo\*idade<sup>2</sup>; Modelo 2 - Considerando-se também a covariável índice de massa corporal (BMI).

## Considerações finais

Os estudos em famílias nucleares permitem avaliar a contribuição das influências genéticas e ambientais para a variação de uma determinada característica, que se expressa numa dada população. Os estudos de agregação familiar e heritabilidade se referem às primeiras duas etapas da EG e possibilitam responder progressivamente às seguintes questões:

a) A variabilidade do fenótipo dentro do seio familiar é menor que entre diferentes famílias? A análise gráfica do comportamento das caixas de extremos e quartis ordenadas pela mediana e o cálculo da ANOVA permitem a resposta a essa primeira pergunta.

b) Se sim, será que os indivíduos com maior grau de parentesco apresentam maior associação do fenótipo em causa? O cálculo dos coeficientes de correlação entre os diferentes pares de familiares é utilizado para responder a essa pergunta.

c) Se há agregação, quanto da  $V_T$  do fenótipo pode ser explicada pelos fatores genéticos? As estimativas de  $h^2$  permitem responder a essa questão.

Em relação aos dados de Muzambinho usados como exemplo neste artigo, a  $h^2$  de C e  $V_{AF}$  foram

significantes, explicando 31 e 24% da VT dos respectivos fenótipos. Esses valores estão de acordo com a variação evidenciada na literatura. Revisões bibliográficas relatam que as  $h^2$  de diferentes fenótipos da composição corporal variam de 30 a 76% (FERMINO et al., 2007), enquanto as do nível de atividade física oscilam de 19 a 57% (CHAVES et al., 2010).

Apesar da grande variação na magnitude de  $h^2$ , a existência de significância em diferentes populações fortalece a presença de um substrato genético importante na determinação daquela característica. Conhecer a magnitude da contribuição específica em cada população permite a elaboração de estratégias também específicas de atuação, o que provavelmente resulta em maior efetividade das mesmas.

Desta forma, continuar estudando diferentes aspectos de similaridade no seio de famílias em diferentes fenótipos da população brasileira, cuja informação seja escassa, é desejável. Esse artigo espera ter estimulado os iniciantes nessa área de investigação, dando-lhes alguma noção do caminho a ser trilhado.

## Notas

1. O valor zero representa o relato, por parte dos entrevistados, de que não se locomoviam com meios que necessitassem de atividade física, sua atividade laboral era basicamente sedentária e não praticavam atividades físicas em seu tempo de lazer.
2. A não se que se especifique com rigor o que se entende por ambiente comum e se possa obter medidas fidedignas desse ambiente.
3. Lembramos os leitores que por ser um estudo de famílias nucleares, a  $h^2$  calculada corresponde ao limite superior de contribuição genética, não sendo possível retirar dessa quantidade a influência do ambiente compartilhado pelos membros de cada família, a não ser que parâmetros objetivos desse ambiente tivessem sido medidos e incluídos no modelo.

## Abstract

Challengers in nuclear family studies: first steps in data analysis

Studies with nuclear families allow the identification of familiar aggregation, as well as the estimation of how much of the variability in a characteristic might be attributed to genetic differences among subjects. Results from these studies are starting points for more complex genetic research, including the identification and relevance of candidate genes. In Sport Science, studies have analyzed the possible existence of familiar aggregation and genetic influence on phenotypes marking physical activity, fitness, and health status. However, few studies were conducted with Portuguese speaking populations, and even fewer with Brazilian populations. Therefore, the present paper aims to discuss, at an introductory level, important aspects about nuclear family studies. The first two phases of Genetic Epidemiology research outline - familiar aggregation and heritability - were approached. The different steps of analysis were

exemplified with real data collected in nuclear families from a study conducted in Muzambinho-MG. We hope this paper provides enough background and stimuli for the newcomers in this area of study.

UNITERMS: Familiar aggregation; Heritability; Waist circumference; Weekly volume of physical activity; Muzambinho.

## Referências

- ALMASY, L.; BLANGERO, J. Multipoint quantitative-trait linkage analysis in general pedigrees. **American Journal of Human Genetics**, Chicago, v.62, n.5, p.1198-211, 1998.
- BASSO, L.; JUNIOR, C.M.; OLIVEIRA, J.A.; FORJAZ, C.L.M.; SOUZA, J.A.D.; PRISTA, A.; MAIA, J.A.R.; TANI, G. Crescimento e desenvolvimento motor de escolares de Muzambinho: um estudo com implicações acadêmicas, sociais e de política interinstitucional. **Revista Portuguesa de Ciências do Desporto**, Porto, v.9, n.2-3, p.247-57, 2009.
- BEUNEN, G.; THOMIS, M. Genetic determinants of sports participation and daily physical activity. **International Journal of Obesity and Related Metabolic Disorders**, London, v.23, p.S55-63, 1999. Supplement 3.
- BRASIL. MINISTÉRIO DA EDUCAÇÃO. INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS. **Censo da Educação Superior. 2007**. Brasília: INEP, 2009. Disponível em: <[http://download.inep.gov.br/download/superior/censo/2007/Resumo\\_tecnico\\_2007.pdf](http://download.inep.gov.br/download/superior/censo/2007/Resumo_tecnico_2007.pdf)>.
- BOUCHARD, C.; MALINA, R.M.; PÉRUSSE, L. **Genetics of fitness and physical performance**. Champaign: Human Kinetics 1997.
- BURTON, P.R.; TOBIN, M.D.; HOPPER, J.L. Key concepts in genetic epidemiology. **Lancet**, New York, v.366, n.9489, p.941-51, 2005.
- CASE WESTERN RESERVE UNIVERSITY. Department of Epidemiology and Biostatistics. **S.A.G.E.: Statistical Analysis for Genetic Epidemiology**. Cleveland: Case Western Reserve University, 2009. Disponível em: <<http://darwin.cwru.edu/sage/>>.
- CHAVES, R.N.; SOUZA, M.C.; SANTOS, D.; GARGANTA, R.; SEABRA, A.; MAIA, J.A.R. Agregação familiar nos níveis de atividade física: um resumo do estado da arte. **Revista Brasileira de Atividade Física e Saúde**, Londrina, v.15, n.1, p.19-26, 2010.
- COLLETTI, G.M.; CARDON, L.R.; FULKER, D.W. A genetic and environmental time series analysis of blood pressure in male twins. **Genetic Epidemiology**, New York, v.10, n.6, p.533-8, 1993.
- DE OLIVEIRA, C.M.; PEREIRA, A.C.; DE ANDRADE, M.; SOLER, J.M.; KRIEGER, J.E. Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. **BMC Medical Genetics**, London, v.9, p.32, 2008.
- FEITOSA, M.F.; KRIEGER, H. O futuro da epidemiologia genética de características complexas. **Ciência e Saúde Coletiva**, Rio de Janeiro, v.7, n.1, p.73-83, 2002.
- FERMINO, R.C.; GARAGANTA, R.; SEABRA, A.; MAIA, J.A.R. Efeitos genéticos e ambientais no indicadores da composição corporal. Uma revisão centrada em estrudos de agregação familiar. **Revista Brasileira de Cineantropometria e Desempenho Humano**, Florianópolis, v.9, n.4, p.414-23, 2007.
- FRIDLEY, B.L.; DE ANDRADE, M. Missing phenotype data imputation in pedigree data analysis. **Genetic Epidemiology**, New York, v.32, n.1, p.52-60, 2008.
- KEEN, K.J.; ELSTON, R.C. Robust asymptotic sampling theory for correlations in pedigrees. **Statistics in Medicine**, Chichester, v.22, n.20, p.3229-47, 2003.
- KRIEGER, H.; MORTON, N.E.; RAO, D.C.; AZEVEDO, E. Familial determinants of blood pressure in northeastern Brazil. **Human Genetics**, Berlin, v.53, n.3, p.415-8, 1980.
- MAIA, J.A.R.; SEABRA, A.; GARGANTA, R.; CHAVES, R.N.; SOUZA, M.C.; SANTOS, D. Factores genéticos e ambientais nos níveis de actividade físico-desportivas: um estudo em famílias nucleares portuguesas. In: BENTO, J.O.; TANI, G.; PRISTA, A. (Ed.). **Desporto e educação física em português**. Porto: Multitema, 2010. p.188-216.
- MAIA, J.A.R.; SEABRA, A.; SILVA, R.M.G.; FREITAS, D.L.; GOUVEIA, E.Q.R.; BASSO, L. Análise de dados gêmeares: uma aventura guiada para investigadores das Ciências do Desporto. **Revista Brasileira de Educação Física e Esporte**, São Paulo, v.25, n.2, p.351-68, 2010.
- PROVINCE, M.A.; RAO, D.C. Path analysis of family resemblance with temporal trends: applications to height, weight, and Quetelet index in northeastern Brazil. **American Journal of Human Genetics**, Chicago, v.37, n.1, p.178-92, 1985.

- RAO, D.C. Genetic dissection of complex traits: an overview. In: RAO, D.C.; PROVINCE, M.A. (Eds.). **Genetic dissection of complex traits**. San Diego: Academic Press, 2001. p.13-34.
- RAO, D.C.; MacLEAN, C.J.; MORTON, N.E.; YEE, S. Analysis of family resemblance. V. Height and weight in northeastern Brazil. **American Journal of Human Genetics**, Chicago, v.27, n.4, p.509-20, 1975.
- RAUDENBUSH, S.; BRYK, A.; CONGDON, R. **HLM6: Hierarchical Linear and Nonlinear Modeling**. Lincolnwood: Scientific Software International, 2004.
- RICE, T.K.; BORECKI, I.B. Familial resemblance and heritability. In: RAO, D.C.; PROVINCE, M.A. (Eds.). **Genetic dissection of complex traits**. San Diego: Academic Press, 2001. p.35-43.
- ROBINSON, W.M.; BORGES-OSORIO, M.R.; CALLEGARI-JACQUES, S.M.; ACHUTTI, A.C.; DA SILVEIRA, L.G.; KLEIN, C.H.; COSTA, E.A. Genetic and nongenetic determinants of blood pressure in a southern Brazilian sample. **Genetic Epidemiology**, New York, v.8, n.1, p.55-67, 1991.
- SOUTHWEST FOUNDATION FOR BIOMEDICAL RESEARCH (SFBR). **SOLAR: Sequential Oligogenic Linkage Analysis Routines**. San Antonio: SFBR, 1995. Disponível em: <<http://www.sfbr.org>>.
- STACORP. **STATA: Data analysis and statistical software**. College Station: StataCorp LP, 1996. Disponível em: <<http://www.stata.com>>.
- UNITED STATES DEPARTMENT OF HEALTH AND HUMAN SERVICES. Centers for Disease Control and Prevention (CDC). **EpiInfo TM 2005**. Atlanta: CDC, 2005. Disponível em: <[www.cdc.gov/epiinfo](http://www.cdc.gov/epiinfo)>.
- VELASQUEZ-MELENDZ, G.; PARRA, F.C.; GAZZINELLI, A.; WILLIAMS-BLANGERO, S.; CORREA-OLIVEIRA, R. Genetic determinants of risk factors for cardiovascular disease in a population from rural Brazil. **Human Biology**, Detroit, v.79, n.2, p.179-90, 2007.
- WIGGINTON, G.E.; ABECASIS, G.R. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. **Bioinformatics**, Oxford, v.21, n.16, p.3445-7, 2005.

## Agradecimentos

Os autores agradecem às famílias que se voluntariaram a participar do estudo de Muzambinho, bem como à Secretaria da Educação e Secretaria da Saúde da cidade, além as diretorias das escolas envolvidas.

Agradecemos ainda, aos alunos e professores que auxiliaram nas coletas de dados. O Setor de Informática da FADE, Universidade do Porto foi essencial para a análise dos dados.

Este estudo recebeu financiamento do CNPq (478249/2007-1), FAPESP (2009/18638-6), Pro Reitorias de Pesquisa (Projeto 1), Extensão (Fundo de Extensão - Projeto Exercício e Coração) e Graduação (Ensinar com Pesquisa 132/2009) da Universidade de São Paulo e FCT (Portugal - PTDC/DES/67569/2006 e FCOMP-01-0124-FEDEB-09608).

ENDEREÇO

Cláudia Lúcia de Moraes Forjaz  
Escola de Educação Física e Esporte - USP  
Av. Prof. Mello Moraes, 65  
05508-030 - São Paulo - SP - BRASIL  
e-mail: cforjaz@usp.br

Recebido para publicação: 06/07/2010

Aceito: 26/01/2011