

O modelo de regressão linear misto para dados longitudinais: uma aplicação na análise de dados antropométricos desbalanceados

Mixed linear regression model for longitudinal data: application to an unbalanced anthropometric data set

Maria Arlene Fausto ¹
 Mariângela Carneiro ²
 Carlos Mauricio de Figueiredo Antunes ²
 Jorge Andrade Pinto ³
 Enrico A. Colosimo ⁴

Abstract

A longitudinal data set is characterized by a time sequence of two or more observations from each individual. In cohort studies, these data are usually not balanced. A data set related to longitudinal height measurements in children of HIV-infected mothers was recorded at the university hospital of the Federal University in Minas Gerais, Brazil. The objective was to assess the application of the mixed effect model to this unbalanced data set. At six months of age, on average boys were 1.8cm taller than girls, and seroreverter infants were 2.9cm taller than their HIV+ peers. At 12 months of age, on average boys were 2.4cm taller than girls and seroreverter children were 3.5cm taller than HIV+ ones. In addition to describing longitudinal height behavior, this model also includes the growth rate estimation for this infant population by gender and group.

Linear Models; Likelihood Functions; Anthropometry; Growth

¹ Escola de Nutrição,
 Universidade Federal de Ouro
 Preto, Ouro Preto, Brasil.

² Instituto de Ciências
 Biológicas, Universidade
 Federal de Minas Gerais,
 Belo Horizonte, Brasil.

³ Faculdade de Medicina,
 Universidade Federal de
 Minas Gerais, Belo Horizonte,
 Brasil.

⁴ Instituto de Ciências
 Exatas, Universidade Federal
 de Minas Gerais, Belo
 Horizonte, Brasil.

Correspondência

E. A. Colosimo
 Departamento de Estatística,
 Instituto de Ciências Exatas,
 Universidade Federal de
 Minas Gerais.
 Av. Antonio Carlos 6627,
 Belo Horizonte, MG
 31270-901, Brasil.
 enricoc@est.ufmg.br

Introdução

Os dados provenientes de estudos longitudinais se caracterizam pela seqüência temporal de duas ou mais observações em cada indivíduo. Nos estudos de coorte, esses dados geralmente apresentam estrutura desbalanceada. O desbalanceamento é uma consequência natural do fato de as medidas obtidas em cada indivíduo serem observadas em tempos diferentes. Além disso, nos estudos de coorte aberta, o tempo total de observação nem sempre é igual para todos os indivíduos, no momento em que os dados vão ser analisados.

Os dados longitudinais apresentam estrutura hierárquica, uma vez que as medidas repetidas são aninhadas dentro do indivíduo ¹. Tal estrutura hierárquica faz com que possamos fazer a suposição de que as observações entre os indivíduos sejam independentes e que as aninhadas no indivíduo possuam a característica da dependência com erros correlacionados. Essa suposição de erros correlacionados exige a modelagem da matriz de co-variância dos dados ^{1,2}. A utilização usual do modelo de regressão linear para esses casos ignora tal correlação porque trata as observações como independentes ³, obtendo inferências menos confiáveis. Em particular, as estimativas dos erros-padrão dos coeficientes do modelo ficam viciadas ^{4,5}. Um modelo estatístico que permite analisar dados longitudinais desbalanceados em estrutura hierárquica, incorpo-

rando a dependência e a estrutura de correlação dos erros é o modelo linear de efeitos mistos ou modelo de efeitos aleatórios ⁶.

Os modelos de efeitos aleatórios para dados longitudinais permitem que os coeficientes da regressão variem entre os indivíduos. Esses modelos têm dois componentes: um intra-individual (uma mudança longitudinal intra-individual é descrita pelo modelo de regressão com um intercepto e inclinação populacional) e outro entre indivíduos (variação no intercepto e inclinação individual).

O modelo de efeitos aleatórios permite não somente descrever a tendência temporal levando em conta a correlação que existe entre medidas sucessivas como também estimar a variação na medida basal e a taxa de mudança ao longo do tempo. No último modelo, as medidas dos indivíduos não precisam ser igualmente espaçadas e balanceadas e as análises podem ser conduzidas com os dados de indivíduos que foram perdidos de seguimento ou que apresentam ausência de informação em algum momento do estudo.

O uso de modelos de efeitos aleatórios é especialmente adequado para dados em que a variabilidade entre os indivíduos é maior do que a variabilidade dentro do indivíduo, como é o caso das curvas de crescimento ⁷. Esse modelo assume que o padrão de crescimento ou de alteração na resposta individual tem a mesma forma funcional para todos os indivíduos, mas que os indivíduos podem apresentar comportamento longitudinal diferente. Isso faz com que cada indivíduo tenha a sua própria curva de crescimento especificada pelos coeficientes da regressão ¹.

Este artigo tem como objetivo demonstrar a aplicação do modelo linear de efeitos mistos na análise de dados de crescimento de crianças.

Avaliação longitudinal do crescimento de lactentes nascidos de mães infectadas com o HIV-1

A avaliação longitudinal do crescimento foi realizada com crianças nascidas de mulheres soropositivas para o HIV-1 (97 lactentes sororrevertores e 42 lactentes vivendo com HIV/AIDS), admitidas nos primeiros três meses de vida para acompanhamento clínico no Ambulatório de AIDS Pediátrica, do Serviço de Doenças Infecciosas e Parasitárias, do Hospital das Clínicas, da Universidade Federal de Minas Gerais (UFMG). A coorte aberta foi iniciada em junho de 1994 e aprovada pelo Comitê Técnico Científico do Hospital das Clínicas da UFMG. A infecção pelo HIV foi determinada com base nos seguintes critérios: existência de duas amostras positivas para

HIV DNA-PCR, pelo método Amplicor Roche ou RNA-PCR, pelo método NASBA Organon-Teknica, após duas semanas de vida; ou, persistência de anticorpos anti-HIV, detectados pelos métodos ELISA e Western Blot, após 18 meses de idade; ou, desenvolvimento de condição clínica definidora de AIDS, em qualquer idade, de acordo com a classificação do Centers for Disease Control and Prevention dos Estados Unidos (CDC) ⁸. Os critérios utilizados para o diagnóstico de sororreversão foram a existência de duas amostras negativas para HIV DNA-PCR, pelo método Amplicor Roche ou RNA-PCR, pelo método NASBA Organon-Teknica, em qualquer idade, após as duas semanas de vida; ou dois testes sorológicos anti-HIV (ELISA) negativos, realizados após os seis meses de idade e com intervalo mínimo de dois meses, na ausência de condições clínicas definidoras de AIDS. O tempo mediano de seguimento dos sororrevertores foi de 14,3 meses, variando de 6,3 a 18,6 meses. O tempo mediano de acompanhamento dos infectados foi de 15,2 meses, variando de 6,8 a 18,0 meses.

Ao utilizar o modelo de efeitos aleatórios, a forma de codificar a variável tempo deve ser levada em consideração no momento de interpretar os resultados provenientes do modelo. Uma das alternativas para a codificação da variável identificadora do tempo, utilizada neste estudo, é aquela em que a medida basal tem o valor 0 (data do nascimento) e as demais medidas aumentam sucessivamente ⁹. Utilizando esse método, o intercepto representa a altura média das crianças no início do estudo.

Os lactentes sororrevertores e infectados contribuíram com 907 e 411 medidas de altura, respectivamente. As medidas antropométricas de altura foram obtidas em cada visita pelo pediatra responsável pelo atendimento. A altura foi aferida com régua antropométrica, em decúbito dorsal.

As co-variáveis incluídas nas análises estatísticas foram: grupo (1 = HIV+; 0 = sororrevertor), sexo (1 = masculino; 0 = feminino) e idade (meses). Entre os lactentes infectados, 47,6% eram do sexo feminino e 78% apresentaram idade gestacional maior ou igual a 36 semanas. Entre os sororrevertores, 45,4% eram do sexo feminino e todos apresentaram idade gestacional a termo. Altura média ao nascimento foi de $48,7 \pm 1,4$ cm entre os sororrevertores e de $48,8 \pm 2,9$ cm entre os infectados, nascidos a termo.

Modelo linear misto

A forma geral do modelo linear misto ¹⁰ é a seguinte:

$$y_{ij} = X'_{ij}\beta + Z'_{ij}\zeta_i + \varepsilon_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n_i, \quad (1)$$

em que: y_{ij} é a resposta da i -ésima criança na j -ésima ocasião, X_{ij} é o vetor de dimensão p de co-variáveis associado aos efeitos fixos β , Z_{ij} é o vetor de dimensão q de co-variáveis associado aos efeitos aleatórios ζ_i e ε_{ij} é o erro aleatório. Assume-se que ζ_i tem distribuição normal com média zero e matriz de variância e co-variância $D(\alpha)$, independente de ε_{ij} que tem distribuição normal com média zero e variância σ^2 .

No modelo linear de efeitos aleatórios, os coeficientes (β) da regressão são denominados efeitos fixos e o conjunto de variâncias e co-variâncias (α), componentes da variância ¹¹.

Cada medida individual difere da média total dos indivíduos em uma determinada quantidade que é definida pelo erro aleatório. Esses erros são independentes e são utilizados para a determinação dos componentes da variância. A co-variância é determinada pelos erros observados na estimativa da média intra-individual e apresentam a característica da dependência.

Os efeitos aleatórios no intercepto representam a heterogeneidade natural entre os indivíduos decorrente de fatores (genéticos ou ambientais) não medidos ^{4,7}. Contudo, parte da variabilidade dos dados também pode ser provocada por diferenças não medidas no nível das co-variáveis, entre os indivíduos e dentro dos indivíduos ². Quando isso acontece, é necessário incluir outro nível de aninhamento no modelo ou, em outras palavras, incluir efeitos aleatórios em outra co-variável. Essa heterogeneidade no nível do indivíduo ou das co-variáveis representa elemento confundidor, que foram omitidos por serem desconhecidos ou por não terem sido avaliados no estudo ¹².

Usualmente, as co-variáveis incluídas no vetor Z são dependentes do tempo e também fazem parte do vetor X . Por exemplo, sexo é uma co-variável que não depende do tempo e, portanto, deve ser incluído em X e não em Z . Por outro lado, contagem de linfócitos T CD₄⁺ é uma co-variável dependente do tempo que é incluída em X e pode ser incluída em Z .

No caso da situação de interesse, somente a co-variável idade é dependente do tempo. Na análise desse estudo foram utilizados dois modelos: um com apenas o intercepto aleatório e o outro com o intercepto e uma inclinação aleatórios. A seguir são apresentados os detalhes desses modelos.

O modelo estatístico com as co-variáveis de interesse e efeitos aleatórios apenas no intercepto é especificado por:

$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{3ij}^2 + \zeta_{0i} + \varepsilon_{ij}$, $i = 1, \dots, m$, $j = 1, \dots, n_i$, em que y_{ij} é a altura da criança na ocasião j , β_0 é o intercepto, ζ_{0i} é o intercepto aleatório, β_1 é o efeito atribuído ao sexo, β_2 é o efeito atribuído à infecção pelo HIV, β_3 e β_4 são os efeitos atribuídos ao polinomial de segundo grau da idade, e ε_{ij} é o erro específico para cada medida que define a quantidade que cada resposta se desvia da trajetória quadrática.

A inclusão da inclinação aleatória faz com que o modelo anteriormente apresentado fique assim especificado:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{3ij}^2 + \zeta_{0i} + \zeta_{3i} x_{3ij} + \varepsilon_{ij}$$

em que: ζ_{3i} permite que a altura de cada criança possa ser diferente da média geral de crescimento na ocasião j . Esses dois modelos são obtidos valendo-se do modelo geral (1).

Nas análises aqui apresentadas, a inclusão de efeitos aleatórios na variável explicativa idade representa a heterogeneidade, entre e dentro dos indivíduos, observada com a modificação da idade. Dessa forma, com a inclusão dos efeitos aleatórios, ajusta-se o modelo por fatores não medidos e obtêm-se estimativas mais precisas do efeito da soroconversão sobre o crescimento.

A estimação dos efeitos fixos (β) e dos componentes da variância (α) do modelo linear misto pode ser feita por meio dos métodos de máxima verossimilhança e máxima verossimilhança restrita ¹³. O método de máxima verossimilhança, ao estimar a variância, assume que não existe erro na estimativa da média e, conseqüentemente, subestima este termo. O método máxima verossimilhança restrita produz estimativas não viciadas da variância, removendo o vício que existe na estimação da média ⁴.

O modelo linear de efeitos mistos foi ajustado por máxima verossimilhança e máxima verossimilhança restrita, utilizando a função LME do S-Plus 2000 (Mathsoft Inc., Seattle, Estados Unidos) ¹⁴.

Etapas da análise

As etapas envolvidas na análise de modelos de efeitos aleatórios são as seguintes: análise exploratória dos dados, ajuste do modelo (efeitos fixos e efeitos aleatórios) e avaliação da adequação do modelo (efeitos aleatórios e termos de interação) ⁹.

A análise exploratória dos dados tem como objetivos avaliar descritivamente a correlação entre as medidas repetidas (A correlação entre as medidas sucessivas é igual? A correlação diminui com o aumento do tempo?), a natureza da tendência temporal (É linear ou não linear?), a heterogeneidade dos dados (Existe variabilidade

de entre as medidas ao longo do tempo? O aumento temporal da variabilidade dos dados torna necessária a inclusão de efeitos aleatórios na inclinação do modelo?) e a presença de valores extremos (*outliers*). As ferramentas utilizadas na análise exploratória são técnicas gráficas e descritivas.

Quando os dados são balanceados (as medidas são obtidas em tempos igualmente espaçados), a análise exploratória pode ser feita por meio da matriz de correlação dos dados (avaliação das correlações das medidas), de gráficos longitudinais de médias (avaliação da tendência temporal), medidas-resumo e análise gráfica da média dos resíduos obtidos com regressão linear ajustada por mínimos quadrados (avaliação da heterogeneidade) e gráficos de perfis (identificação de valores extremos). Quando se trabalha com dados temporalmente desbalanceados (as medidas são obtidas em tempos diferentes), a análise exploratória dos dados para avaliar a tendência temporal e a heterogeneidade dos dados pode ser realizada com ferramentas gráficas que forneçam médias suavizadas. Com dados desbalanceados, a avaliação da correlação dos dados por meio de matriz de correlação só pode ser realizada com a utilização de uma variável categórica do tempo.

A avaliação da estrutura de correlação dos dados é importante quando os dados são balanceados e se pretende trabalhar com um modelo marginal para a análise dos dados. A principal característica desse modelo é que a média e a variância da resposta investigada são modeladas separadamente. As principais formas de modelagem da estrutura de variância e co-variância utilizam matriz de correlação não estruturada ou correlação uniforme ou correlação auto-regressiva. O modelo auto-regressivo exige que as medidas sejam obtidas em tempos igualmente espaçados, não sendo aplicável para dados desbalanceados¹⁵. Dados desbalanceados e com um número grande de observações para cada indivíduo também não permitem que se trabalhe com uma matriz de correlação não estruturada⁶. A modelagem utilizando a correlação uniforme trabalha com a suposição de que a correlação entre as medidas de todos os indivíduos é constante ao longo do tempo. Trabalhar com essa suposição quando se modelam curvas de crescimento significa negar que influências biológicas e ambientais podem fazer os indivíduos apresentarem taxas de crescimento diferentes num mesmo estágio de vida. Por esse motivo, o modelo marginal não é adequado nesta situação¹³. Ao utilizar o modelo de efeitos aleatórios, é assumido que existe correlação entre as medidas repetidas e que cada indivíduo tem um nível não observado

de resposta que persiste por intermédio das medidas, em função do tempo¹⁵.

A tendência temporal dos dados e o comportamento longitudinal da altura dos lactentes sororvertentes e infectados pelo HIV foram avaliados por meio de análise gráfica. As curvas de crescimento foram suavizadas pelo método Lowess¹⁶.

A Figura 1 mostra que o crescimento da altura apresenta um comportamento não linear. Esse comportamento torna necessário que o modelo seja ajustado com um termo quadrático na variável idade. A distribuição dos dados também mostra que alguns lactentes são maiores ao nascimento do que os outros, tornando necessária a inclusão do intercepto aleatório.

A Figura 2 mostra que, ao longo do tempo, os lactentes sororvertentes parecem ser maiores do que os infectados.

A Figura 3 mostra que os meninos parecem ser maiores do que as meninas.

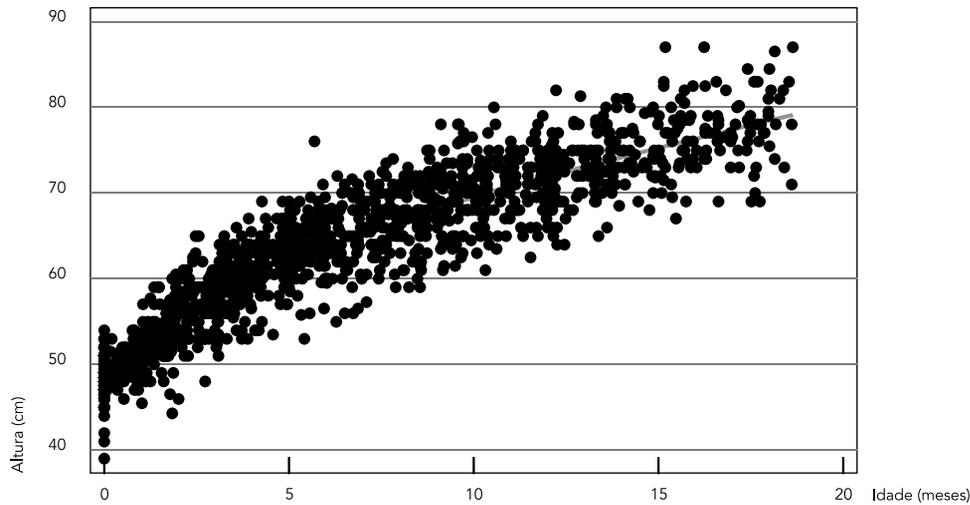
Estimativas utilizando o método de máxima verossimilhança e máxima verossimilhança restrita

Na Tabela 1 são apresentados os resultados da análise univariada que utiliza a função LME do S-Plus e os estimadores de máxima verossimilhança e máxima verossimilhança restrita. As variáveis significativas para o ajuste do modelo de altura foram sexo e o polinomial da idade. A interpretação dos coeficientes (β) é exatamente a mesma dos modelos para dados transversais. A análise univariada indica que os meninos são, em média, 1,44cm mais altos que as meninas. O polinomial da idade mostra que a altura aumenta com a idade e a presença de um efeito negativo no polinomial demonstra que a velocidade de crescimento da altura diminui com o aumento da idade. A variável grupo foi incluída na análise multivariada para avaliar se existiam diferenças no crescimento de altura de lactentes infectados com o HIV-1 e sororvertentes. O efeito da variável grupo não foi estatisticamente significativo. Não foram observadas diferenças nas estimativas dos componentes de variância que usam os estimadores de máxima verossimilhança e máxima verossimilhança restrita (valores não mostrados). O valor p para os testes das respectivas hipóteses é o da estatística de Wald¹³.

Na Tabela 2, são apresentados os ajustes para a análise multivariada utilizando o modelo de regressão linear de efeitos mistos para a altura. Todas as variáveis foram significativamente associadas com a altura. O intercepto do modelo mostra que a altura média ao nascimento das

Figura 1

Distribuição longitudinal da altura das crianças nascidas de mães vivendo com HIV/AIDS. Coorte de Belo Horizonte, Minas Gerais, Brasil, 1995-2004.



crianças nascidas de mães infectadas pelo HIV-1 era de 48,9cm. Ao somar-se o efeito do intercepto com o da variável sexo, observa-se que a média de 48,9cm era aplicável às crianças do sexo feminino e que a altura média ao nascimento dos meninos era de 50,45cm. Quando se somam os efeitos do intercepto, da variável sexo e da variável grupo, obtêm-se as seguintes informações sobre a altura média ao nascimento dessas crianças por sexo e grupo: (1) as crianças do sexo feminino não infectadas nasceram com altura média de 48,9cm e as infectadas com 46,38cm; (2) as crianças do sexo masculino não infectadas nasceram com altura média de 50,45cm e as infectadas com 47,93cm. O polinomial da idade revela que a altura aumenta com a idade, e a presença de um efeito negativo no polinomial indica que a velocidade de crescimento da altura diminui com o aumento da idade.

Durante a construção dos modelos, quando as variáveis sexo e polinomial da idade foram associados, ocorreu uma leve alteração do coeficiente da variável sexo, e quando se associaram a variável grupo e o polinomial da idade, ocorreu modificação de efeito na estimativa da variável grupo, se comparados com os valores obtidos na análise univariada (Tabela 1).

Ao comparar os modelos ajustados por máxima verossimilhança e máxima verossimilhança restrita, apresentados na Tabela 2, observa-se que as diferenças na estimativa do desvio-pa-

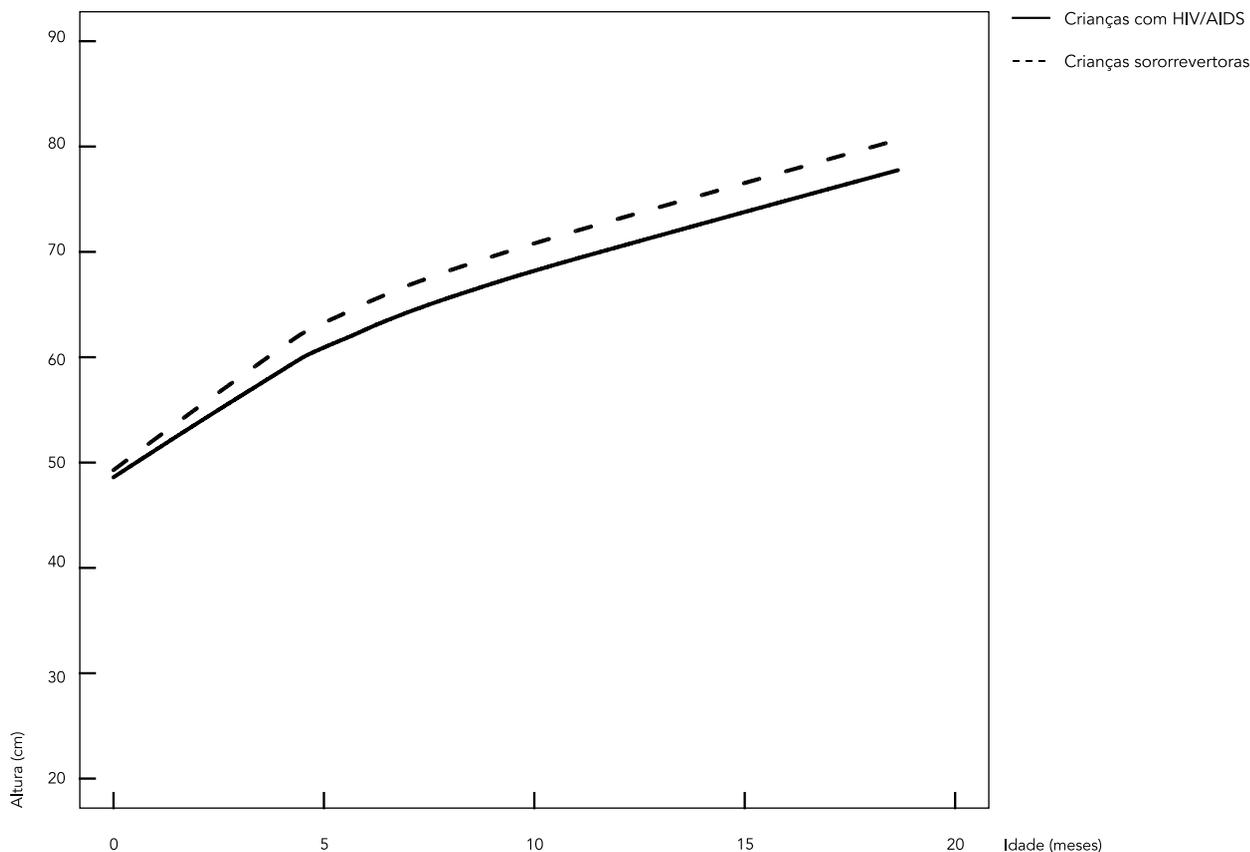
drão de algumas variáveis aparecem na terceira casa decimal e que há um leve aumento na estimativa do desvio-padrão do intercepto aleatório, indicando provavelmente que o modelo ficou mais bem ajustado com máxima verossimilhança restrita.

Na Tabela 3, é apresentada a análise multivariada que utiliza o modelo de regressão linear de efeitos mistos para estimar a altura, ajustados com máxima verossimilhança restrita, com efeitos aleatórios no intercepto e na idade. Com a inclusão de efeitos aleatórios na variável idade, verificou-se que houve modificação no valor da estimativa da média para a variável grupo de -2,52kg (modelo 2) para -1,99kg. Também foi observado que ocorreu uma leve modificação no coeficiente da variável sexo. Essas modificações de efeito sugerem a existência de interação entre as variáveis grupo e sexo e o polinomial da idade.

Na Tabela 4 são apresentados os modelos lineares mistos finais para a altura, ajustados com máxima verossimilhança e máxima verossimilhança restrita, com efeitos aleatórios no intercepto e na variável idade. A significância dos efeitos aleatórios na variável idade foi avaliada pelo teste da razão de máxima verossimilhança que indicou que a inclusão do efeito aleatório na inclinação da curva de crescimento melhorou o ajuste do modelo (Tabela 5). Tal teste somente é válido se o modelo apresentado na Tabela 4

Figura 2

Curvas médias suavizadas de crescimento da altura de crianças com HIV/AIDS e sororrevertoras. Coorte de Belo Horizonte, Minas Gerais, Brasil, 1995-2004.



for adequado para os dados. A adequação do modelo é realizada essencialmente por meio dos resíduos do modelo ajustado. As Figuras 4, 5 e 6 mostram gráficos de adequação. A Figura 4 mostra que os valores ajustados estão bem próximos dos observados. As Figuras 5 e 6 mostram que as suposições de normalidade e homocedasticidade dos erros parecem atendidas pelos dados.

No modelo da altura ajustado por máxima verossimilhança, a inclusão de efeitos aleatórios na variável idade também foi estatisticamente significativo para o ajuste do modelo ($p < 0,0001$).

A indicação de existência de interações no modelo para estimar a altura foi confirmada nos modelos ajustados por máxima verossimilhança e máxima verossimilhança restrita (Tabela 4). As estimativas da média e do desvio-padrão para as variáveis do modelo de altura, utilizando os

dois estimadores, foram similares. Foi verificada apenas uma leve diferença no desvio-padrão do intercepto aleatório quando se comparou o modelo ajustado por máxima verossimilhança e o ajustado por máxima verossimilhança restrita.

Os modelos finais apresentados na Tabela 4 confirmam as informações observadas na análise gráfica descritiva que indicavam uma diferença maior na altura dos meninos quando comparados com as meninas com o aumento da idade e que as diferenças na altura de crianças sororrevertoras e infectadas aumentavam com a idade. As equações que descrevem o crescimento em centímetros das crianças são as seguintes:

1) Altura (crianças sororrevertoras) = $49 + (0,58 \times \text{sexo}) + (2,98 \times \text{idade}) - (0,08 \times \text{idade}^2) + (0,26 \times \text{sexo} \times \text{idade}) - (0,009 \times \text{sexo} \times \text{idade}^2)$

Figura 3

Curvas médias suavizadas de crescimento da altura de crianças com HIV/AIDS e sororrevertoras, por sexo. Coorte de Belo Horizonte, Minas Gerais, Brasil, 1995-2004.

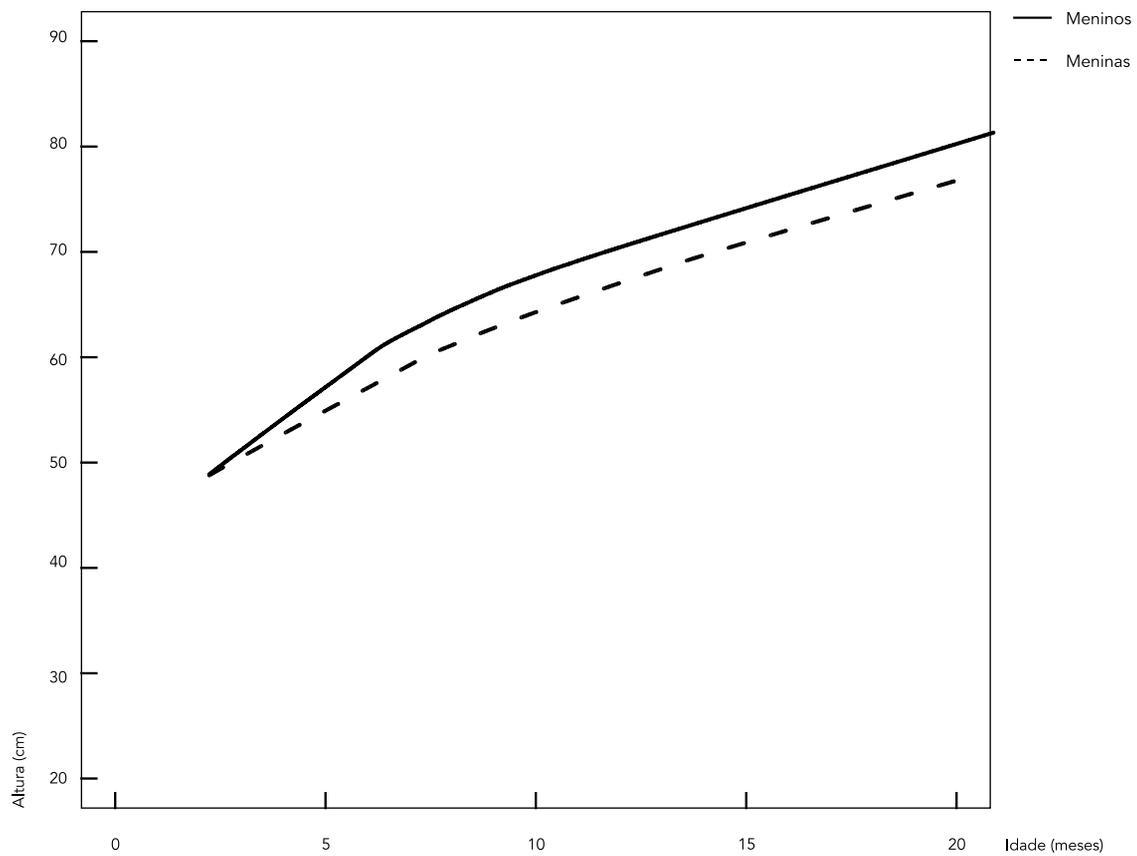


Tabela 1

Análise univariada utilizando o modelo de regressão linear de efeitos mistos com intercepto aleatório para o modelo de altura das crianças infectadas e sororrevertoras, utilizando os métodos de máxima verossimilhança e máxima verossimilhança restrita.

	Máxima verossimilhança		Máxima verossimilhança restrita	
	β	p	β	p
Sexo	1,44	0,02	1,44	0,02
Grupo	0,25	0,69	0,25	0,69
Polinomial				
Idade	2,98	< 0,0001	2,98	< 0,0001
Idade ²	-0,08	< 0,0001	-0,08	< 0,0001

Tabela 2

Análise multivariada da regressão linear de efeitos mistos para o modelo de altura das crianças infectadas e sororrevertoras, utilizando o método de máxima verossimilhança e máxima verossimilhança restrita e efeitos aleatórios no intercepto.

	Máxima verossimilhança			Máxima verossimilhança restrita		
	β	DP	p	β	DP	p
Intercepto	48,90	0,376	< 0,0001	48,90	0,379	< 0,0001
Sexo	1,55	0,446	0,0007	1,55	0,449	0,0008
Grupo	-2,52	0,485	< 0,0001	-2,52	0,489	< 0,0001
Idade	2,97	0,038	< 0,0001	2,97	0,038	< 0,0001
Idade ²	-0,08	0,002	< 0,0001	-0,08	0,002	< 0,0001
Efeitos aleatórios						
Intercepto	2,51255			2,54245		
Resíduos	2,06821			2,06994		
Observações (n)	1.318			1.318		
Crianças (n)	139			139		

Tabela 3

Análise multivariada da regressão linear de efeitos mistos para o modelo de altura das crianças infectadas e sororrevertoras, utilizando o método de máxima verossimilhança restrita, com efeitos aleatórios no intercepto e na inclinação.

	β	DP	p
Intercepto	48,9634	0,34595	< 0,0001
Sexo	1,19836	0,41572	0,0046
Grupo	-1,99432	0,45526	< 0,0001
Idade	2,98418	0,03898	< 0,0001
Idade ²	-0,08006	0,00212	< 0,0001
Efeitos aleatórios			
Intercepto	2,32333		
Idade	0,23688		
Resíduos	1,73063		
Observações (n)	1.318		
Crianças (n)	139		

2) Altura (crianças HIV+) = $49 + (0,58 \times \text{sexo}) - 0,88 + (2,98 \times \text{idade}) - (0,08 \times \text{idade}^2) + (0,26 \times \text{sexo} \times \text{idade}) - (0,009 \times \text{sexo} \times \text{idade}^2) - (0,46 \times \text{idade}) + (0,02 \times \text{idade}^2)$

Com base nessas equações, pode-se observar que, aos seis meses de idade, os meninos eram, em média, 1,8cm maiores que as meninas e as crianças sororrevertoras eram, em média, 2,9cm maiores que as infectadas. Aos 12 meses, a diferença na altura média entre infectados e sororrevertores passou a ser de 3,5cm enquanto entre os meninos e as meninas passou a ser, em média, de 2,4cm. Essas equações também podem ser utilizadas para estimar a velocidade de crescimento

das crianças infectadas e sororrevertoras, por sexo, pelas seguintes equações (derivadas matemáticas do modelo longitudinal).

- **Sexo feminino**

3) Velocidade de crescimento (crianças sororrevertoras) = $2,98 - (0,08 \times \text{idade})$

4) Velocidade de crescimento (crianças HIV+) = $2,98 - (0,08 \times \text{idade}) - 0,46 + (0,02 \times \text{idade})$

Tabela 4

Modelo para estimar a altura das crianças infectadas e sororrevertoras ajustados por máxima verossimilhança e máxima verossimilhança restrita.

	Máxima verossimilhança restrita				Máxima verossimilhança			
	β	DP	IC95%	p	β	DP	IC95%	p
Intercepto	49,00	0,36	48,30; 49,72	< 0,0001	49,00	0,36	48,30; 49,71	< 0,0001
Sexo	0,58	0,45	-0,31; 1,47	0,20	0,58	0,45	-0,30; 1,46	0,27
Grupo	-0,88	0,50	-1,86; 0,10	0,08	-0,88	0,49	-1,86; 0,09	0,04
Idade	2,98	0,06	2,86; 3,10	< 0,0001	2,98	0,06	2,86; 3,10	< 0,0001
Idade ²	-0,08	0,003	-0,09; -0,07	< 0,0001	-0,08	0,003	-0,09; -0,07	< 0,0001
Interação								
Sexo*idade	0,26	0,08	0,11; 0,41	0,0007	0,26	0,08	0,11; 0,41	< 0,0001
Sexo*idade ²	-0,009	0,004	-0,02; -0,001	0,03	-0,009	0,004	-0,02; -0,001	0,001
Grupo*idade	-0,46	0,08	-0,63; -0,30	< 0,0001	-0,46	0,08	-0,63; -0,30	< 0,0001
Grupo*idade ²	0,02	0,004	0,01; 0,03	< 0,0001	0,02	0,004	0,01; 0,03	< 0,0001
Efeitos aleatórios								
Intercepto	2,32		1,89; 2,86		2,29		1,86; 2,81	
Idade	0,22		0,18; 0,28		0,22		0,18; 0,28	
Resíduos	1,71		1,61; 1,82		1,70		1,60; 1,81	
Observações (n)	1.318				1.318			
Crianças (n)	139				139			

Tabela 5

Teste da razão de máxima verossimilhança para avaliar a inclusão dos efeitos aleatórios no modelo da altura.

Modelo	Log-Likelihood	Teste	p
1	-3021.369	-	-
2	-2904.080	1 vs. 2	< 0,0001

• Sexo masculino

5) Velocidade de crescimento (crianças sororrevertoras) = $2,98 - (0,08 \times \text{idade}) + 0,26 - (0,009 \times \text{idade})$

6) Velocidade de crescimento (crianças HIV+) = $2,98 - (0,08 \times \text{idade}) + 0,26 - (0,009 \times \text{idade}) - 0,46 + (0,02 \times \text{idade})$

Utilizando essas equações, pode-se estimar que aos seis meses de idade a velocidade média de crescimento das crianças sororrevertoras do sexo masculino e do sexo feminino era de 32,5 e 30cm/ano e que a velocidade média de crescimento das crianças infectadas do sexo masculino e do sexo feminino era de 28,4 e 25,9cm/ano. Aos 12 meses, a velocidade média de crescimento das crianças sororrevertoras do sexo masculino e do sexo feminino era de 26,1 e 24,2cm/ano e que a velocidade média de crescimento das crianças

infectadas do sexo masculino e do sexo feminino era de 23,1 e 21,6cm/ano.

Discussão

Os indicadores antropométricos, dentre outros objetivos, são utilizados para identificar indivíduos em risco nutricional e avaliar a efetividade de intervenções¹⁷. O acompanhamento longitudinal para monitorar o crescimento de crianças exige a adoção de técnicas estatísticas mais sofisticadas que permitam contornar problemas inerentes à coleta das informações que tornam os dados desbalanceados e levar em conta a estrutura de dependência existente entre as medidas.

Ao se ignorar a existência de medidas repetidas e ao utilizar um modelo de regressão linear ajustado por mínimos quadrados, ocorre pouca

Figura 4

Gráficos de valores ajustados versus observados para o modelo da Tabela 4.

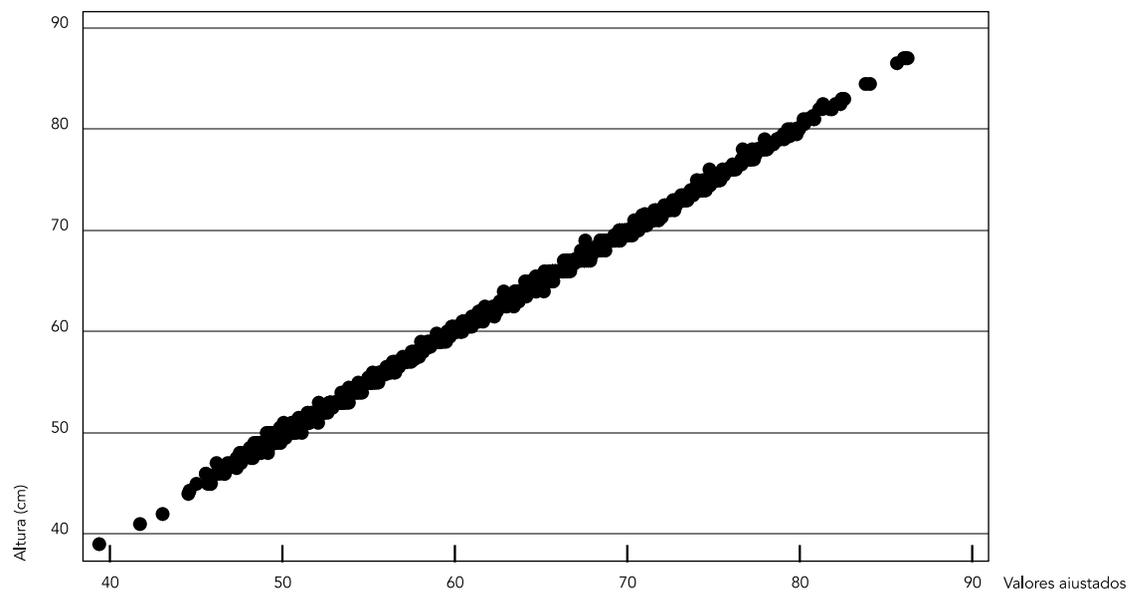


Figura 5

Gráfico de probabilidade normal para os resíduos do modelo da Tabela 4.

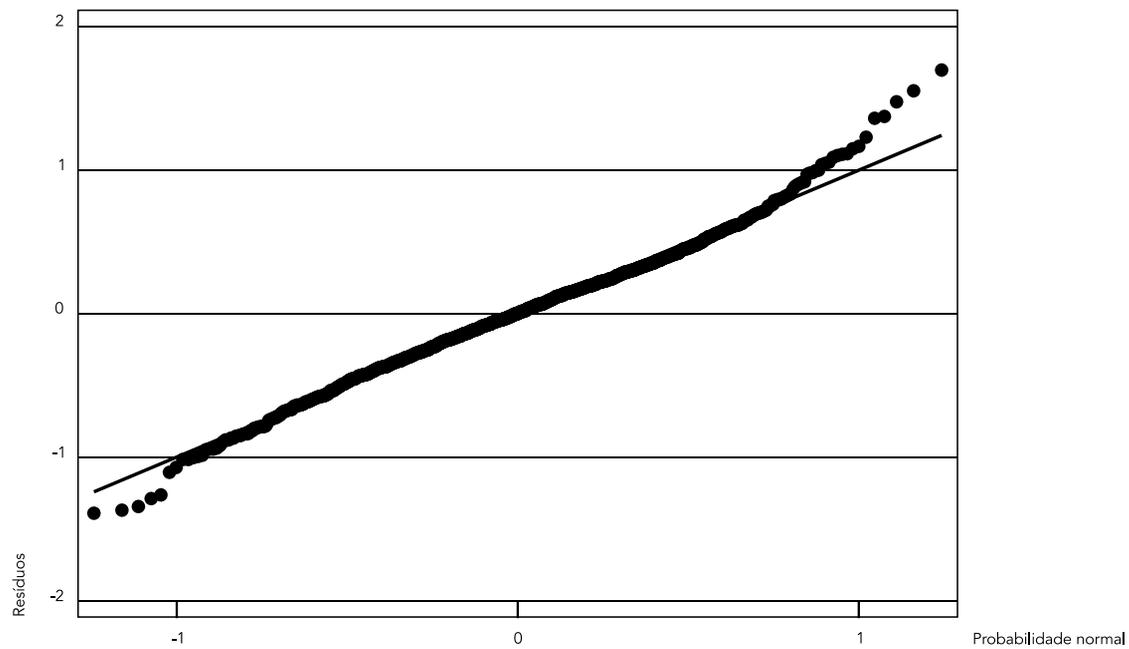
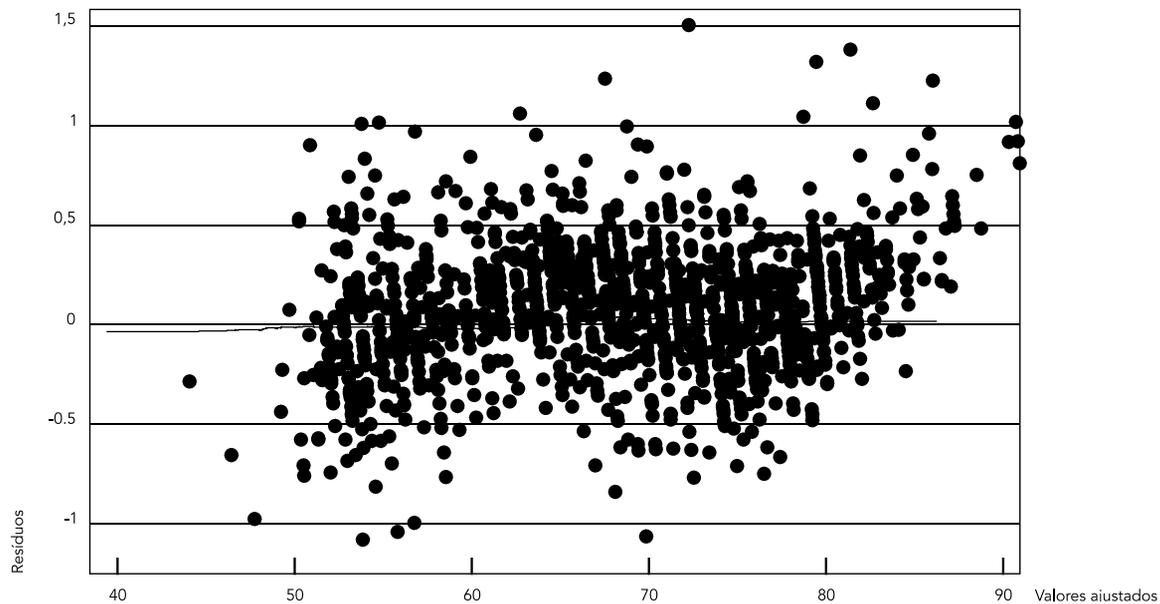


Figura 6

Gráfico de valores ajustados versus resíduos para o modelo da Tabela 4.



diferença na estimativa da média e aumento do erro na estimativa do desvio-padrão. Esse vício pode resultar na não-deteção de diferenças no crescimento em co-variáveis investigadas. O uso do modelo linear de efeitos mistos permite contornar esses problemas e, por conseguinte, modelar de forma adequada os dados correlacionados.

Os resultados mostram diferenças entre as alturas médias de meninos e meninas e também das crianças sororrevertoras e infectadas. As diferenças em ambos os casos aumentam com a idade. As velocidades de crescimento das crianças sororrevertoras e infectadas de ambos os sexos diminuem com a idade.

A provável razão para não se ter observado diferença nas estimativas dos componentes da variância utilizando os estimadores de máxima verossimilhança e máxima verossimilhança restrita é o fato de ter se trabalhado com uma amostra de grande tamanho. Na presença de amostras de pequeno tamanho, ou se estiverem sendo investigadas muitas co-variáveis simultaneamente, o estimador de máxima verossimilhança restrita é o mais adequado para o ajuste de modelos lineares mistos.

O modelo linear de efeitos mistos mostrou-se adequado para analisar estudos longitudinais

com dados desbalanceados. A aplicabilidade desse modelo em Saúde Pública pode ser estendida a diferentes áreas, podendo entre elas, citar: avaliações de resposta imunológica e de testes laboratoriais em indivíduos infectados ou doentes, acompanhados em estudo de coorte (nos pacientes HIV positivos, as medidas de CD4, carga viral, hemograma, entre outras), avaliações das infecções respiratórias e diarreicas e resposta de fase aguda em crianças acompanhadas por um período de tempo, entre outras. No caso das curvas de crescimento, o modelo permite não apenas descrever o comportamento longitudinal das variáveis investigadas, mas também estimar as velocidades de crescimento.

Resumo

Os dados provenientes de estudos longitudinais se caracterizam pela seqüência de duas ou mais observações em cada indivíduo. Nos estudos de coorte, esses dados geralmente apresentam estrutura desbalanceada. Uma casuística que envolve a avaliação longitudinal de crescimento de lactentes nascidos de mães infectadas pelo HIV foi acompanhada no ambulatório de AIDS pediátrica do Hospital das Clínicas da Universidade Federal de Minas Gerais, Minas Gerais, Brasil. O objetivo deste estudo é demonstrar a aplicação do modelo linear misto na análise de dados longitudinais desbalanceados provenientes dessa coorte. Os resultados mostram que, aos seis meses de idade, os meninos eram, em média, 1,8cm maiores que as meninas e as crianças sororrevertoras eram, em média, 2,9cm maiores que as infectadas. Aos 12 meses, a diferença na altura entre meninos e meninas passou a ser, em média, de 2,4cm enquanto a diferença entre infectados e sororrevertores passou a ser, em média, de 3,5cm. Além de descrever o comportamento longitudinal do crescimento, o modelo também permite estimar a velocidade de crescimento das crianças por sexo e grupo.

Modelos Lineares; Funções Verossimilhança; Antropometria; Crescimento

Referências

1. Ker H-W, Wardrop J, Anderson C. Application of linear mixed-effects models in longitudinal data: a case study. http://www.hiceducation.org/Edu_Proceedings/Hsiang-Wei%20Ker.pdf (acessado em 30/Jul/2003).
2. Pinheiro JC, Bates DM. LME and nLME: mixed effects models methods and classes for S and S-Plus. Version 1.2. Madison: University of Wisconsin-Madison; 1995.
3. Draper N, Smith H. Applied regression analysis. 3rd Ed. New York: John Wiley and Sons; 1998. (Wiley Series in Probability and Statistics).
4. Diggle PJ, Liang K-Y, Zeger SL. Analysis of longitudinal data. Oxford: Clarendon Press; 1994.
5. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73:13-22.
6. Cnnan A, Laird NM, Slasor P. Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Stat Med* 1997; 16:2349-80.
7. Molenberghs G, Verbeke GA. Review on linear mixed models for longitudinal data, possibly subject to dropout. http://www.biostatistik.uibk.ac.at/roes/papers/RoES_2001_Molenberghs_Verbeke_Paper.pdf (acessado em 30/Jul/2003).
8. Centers for Disease Control and Prevention. 1994 revised classification system for human immunodeficiency virus infection in children less than 13 years of age. *MMWR Morb Mortal Wkly Rep* 1994; 43:1-13.
9. Ekuma O, Lix L. Random effects models for longitudinal data – continuous data. Manitoba: Manitoba Centre for Health Policy; 2004.
10. Laird NM, Ware SH. Random-effects models for longitudinal data. *Biometrics* 1982; 38:963-74.
11. Goldstein H, Browne W, Rabash J. Multilevel modelling of medical data. *Stat Med* 2002; 21:3291-315.
12. Skrondal A, Rabe-Hesketh S. Some applications of generalized linear latent and mixed models in epidemiology: repeated measures, measurement error and multilevel modeling. *Norwegian Journal of Epidemiology* 2003; 13:265-78.
13. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer-Verlag; 2000. (Springer Series in Statistics).
14. Pinheiro JC, Bates DM. Mixed effects models in S and S-Plus. New York: Springer-Verlag; 2000. (Statistics and Computing).
15. Belloc R. Analysis of longitudinal data in Stata, S-Plus and SAS. Stockholm: Department of Medical Epidemiology, Karolinska Institutet; 2001.
16. Royston P. Lowess smothing. *Stata Technical Bulletin* 1991; 3:7-9.
17. World Health Organization. Physical status: the use and interpretation of anthropometry. Geneva: World Health Organization; 1995. (WHO Technical Report Series, 854).

Colaboradores

J. A. Pinto é o pesquisador responsável pela coorte de crianças infectadas pelo HIV. M. A. Fausto e E. A. Colosimo pela realização da análise estatística e interpretação dos resultados. M. Carneiro e C. M. F. Antunes contribuíram no desenho do estudo epidemiológico.

Agradecimentos

Os autores agradecem à Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento do projeto. Os autores também agradecem a um revisor as sugestões que melhoraram a presente versão do artigo.

Recebido em 05/Jun/2006
Versão final reapresentada em 29/Ago/2007
Aprovado em 03/Set/2007