

**CURVA DE CRESCIMENTO USANDO MODELO MISTO:
UMA APLICAÇÃO NA PROGRESSÃO DA DOENÇA DE MACHADO-JOSEPH**

*GROWTH CURVE USING MIXED MODEL: AN APPLICATION ON
MACHADO-JOSEPH DISEASE PROGRESSION*

Lisiane Hauser, Álvaro Vigo^{1,5}, Christian Kieling², Laura Bannach Jardim^{2,3},
Suzi Alves Camey^{1,4,5}, Vanessa Bielefeldt Leotti¹

RESUMO

Introdução: A obtenção dos dados por meio de medidas repetidas em diversas ocasiões no tempo em um mesmo sujeito torna possível o ajuste de curvas que descrevam padrões de evolução e identificam preditores de evolução. O objetivo deste trabalho foi ajustar curvas para descrever a progressão da doença de Machado-Joseph (DMJ), quantificada pelo escore NESSCA (*Neurological Examination Score for Spinocerebellar Ataxia*), utilizando modelos mistos.

Métodos: Os dados foram obtidos de uma coorte de pacientes da DMJ acompanhada no Hospital de Clínicas de Porto Alegre (HCPA) durante um período de 10 anos. Nas avaliações clínicas realizadas, o comprometimento clínico do paciente foi mensurado várias vezes através do escore NESSCA. Esse escore foi considerado como desfecho, e a progressão da doença poderia ser influenciada pelas variáveis explicativas: idade no início da doença e comprimento da mutação. O procedimento *Proc MIXED* do software SAS foi utilizado para realizar o ajuste dos modelos.

Resultados: Progressão da doença ocorre mais lentamente com o aumento na idade de início da doença, por outro lado, com o aumento do comprimento da mutação, mais rápida é a progressão da doença.

Conclusão: Uma maior idade no início da doença é fator de proteção para a progressão da DMJ e um maior comprimento da mutação é fator de risco. Ressalta-se que as atribuições de proteção e risco estão relacionadas exclusivamente com a velocidade de progressão da doença, não sendo observados efeitos significativos dessas variáveis para o escore no início da doença.

Unitermos: *Curvas de crescimento; modelos mistos; doença de Machado-Joseph; NESSCA*

ABSTRACT

Background: Obtaining data with repeated measures in different occasions, from the same patient, makes the growth curve adjustment possible. These curves describe evolution patterns and identify evolution predictors. The main purpose of this study was to adjust growth curve to describe Machado-Joseph disease progression (MJD), quantified by the NESSCA score (*Neurological Examination Score for Spinocerebellar Ataxia*) using linear mixed model.

Methods: the data were obtained from a cohort of MJD patients observed at Hospital de Clínicas de Porto Alegre (HCPA) during 10 years. In the clinical evaluation performed, the clinical implications of these patients were measured by the NESSCA score. In order to accomplish the data analysis, the NESSCA score was used as an outcome and it was considered that the disease progression could be influenced by explaining variables such as age at onset and CAG length. The procedure *Proc MIXED* of the software SAS was used to perform the models adjustment.

Results: the disease progression is slower with the increase of age at onset, on the other hand, the progression of the disease is faster with CAG length increase.

Conclusion: a higher age at onset is a protection factor to the MJD progression and a higher CAG length is a risk factor. It is highlighted that the attributions of protection and risk are exclusively related with the progression speed of the disease, since there were no significant effects of these variables to the score at the beginning of the disease.

Keywords: *Growth curve; mixed model; Machado-Joseph disease; NESSCA*

Rev HCPA 2009;29(1):05-17

Dados longitudinais estão presentes em vários estudos na área da saúde. Esses dados são medidas repetidas em diversas ocasiões no tempo em uma mesma unidade experimental (sujeitos, indivíduos ou objetos) com o objetivo de verificar o seu comportamento ao longo do tempo. Em especial, em muitos estudos desse tipo, quando o desfecho é contínuo, deseja-se ajustar curvas que descrevam padrões de evolução e identifiquem preditores de evolução. O

ordenamento temporal das observações é importante, porque observações dentro de um sujeito, em um tempo próximo, têm maior probabilidade de serem similares do que aquelas separadas por um tempo maior.

No caso de o intervalo entre dois tempos quaisquer consecutivos serem constante ao longo do estudo, os dados longitudinais são denominados regulares. Se as observações são feitas nos mesmos instantes de tempo para

1. Departamento de Estatística, Instituto de Matemática, Universidade Federal do Rio Grande do Sul (UFRGS).

2. Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre (HCPA).

3. Departamento de Medicina Interna, Faculdade de Medicina, UFRGS.

4. Grupo de Pesquisa e Pós-Graduação, HCPA.

5. Programa de Pós-graduação em Epidemiologia, UFRGS.

Contato: Susy Camey, e-mail: camey@mat.ufrgs.br, (Porto Alegre, RS, Brasil).

todos os sujeitos, tem-se uma estrutura de dados balanceada e, além disso, a ausência de observações perdidas caracteriza uma estrutura de dados completa (1).

Os dados longitudinais apresentam uma hierarquia natural, também chamada de estrutura multinível, uma vez que as medidas repetidas são aninhadas dentro do sujeito. Tal hierarquia faz com que possamos fazer a suposição de que as observações entre os sujeitos sejam independentes e que as aninhadas no sujeito sejam dependentes/correlacionadas. Essa existência de correlação exige a modelagem da matriz de covariância dos dados.

O ajuste de curvas de crescimento pode ser realizado através da análise de perfis utilizando um modelo univariado, de acordo com um delineamento de parcelas subdivididas (*split plot design*), que impõe forte restrição quanto à matriz de variâncias-covariâncias. Alternativamente, pode ser usado um modelo multivariado, que utiliza uma matriz de variâncias-covariâncias sem restrições, chamada não-estruturada (1,2). O método de análise do modelo univariado assume diferenças individuais para os interceptos, mas não assume tais diferenças para as inclinações das curvas que descrevem o comportamento dos dados ao longo do tempo. Dessa forma, quando há diferenças individuais nas inclinações, o método subestima a variabilidade e superestima os resultados dos testes estatísticos. Além disso, não considera estrutura de dados desbalanceada, ou seja, o método desconsidera um sujeito no caso de existir uma ou mais observações faltantes. Já o método multivariado considera as diferenças nas inclinações, isto é, curvas separadas são estimadas para cada sujeito e os parâmetros das curvas individuais são utilizados para estimar as curvas dos grupos. Entretanto, o método não incorpora com facilidade covariáveis que variam ao longo do tempo, desconsidera dados não-regulares e, assim como no método univariado, não leva em conta estrutura de dados desbalanceada (3).

Outra forma de ajustar curvas de crescimento é através de modelos mistos lineares ou não-lineares, que possibilitam o uso de diferentes tipos de estrutura para as matrizes de variâncias-covariâncias, optando-se por aquela que melhor representa a estrutura de correlação dos dados, e também permite descrever o comportamento dos perfis médios através de curvas (1). O modelo misto considera a existência de efeitos fixos e aleatórios, onde os efeitos são considerados fixos quando se deseja inferir somente para as categorias das variáveis independentes que são incluídas no ajuste do modelo, pois essas representam as categorias sobre as quais se deseja fazer inferências. Os efeitos são considerados aleatórios quando as categorias das variáveis utilizadas no ajuste do modelo constituem uma amostra aleatória de uma população,

ou seja, um conjunto de categorias com uma distribuição de probabilidade, e se deseja inferir para a população inteira. Neste caso, os efeitos fixos são constantes ao longo das unidades experimentais (sujeitos) e efeitos aleatórios variam (4). Este modelo pode ser especificado na forma de um modelo hierárquico, sendo o primeiro nível usado para modelar a variabilidade dentro dos sujeitos e o segundo nível para modelar a variabilidade entre os sujeitos (5).

Há dois tipos de modelos mistos úteis para a análise de dados longitudinais: o modelo Padrão de Covariância e modelo de Coeficientes Aleatórios. O modelo Padrão de Covariância é utilizado quando os dados são balanceados; em contrapartida o modelo de Coeficientes Aleatórios é apropriado para dados desbalanceados ou se há interesse particular na relação entre a variável resposta e o tempo. Neste caso, curvas de regressão são ajustadas para cada sujeito e são admitidos coeficientes de regressão variando aleatoriamente entre os sujeitos. Essa variação ocorre, de forma simples, nos interceptos ou, de forma mais complexa, também nas inclinações (6).

Os modelos mistos são também conhecidos como modelos lineares hierárquicos e modelos lineares multiníveis, terminologias especialmente comuns na literatura educacional, nas ciências sociais, na medicina e na epidemiologia, entre outras.

Neste trabalho, foram estimadas curvas de crescimento da Doença de Machado-Joseph (DMJ). Esta doença, também conhecida como ataxia spinocerebellar tipo 3 (SCA3), é uma desordem neurodegenerativa autossômica dominante caracterizada pela ataxia cerebelar progressiva que acarreta, sem nunca alterar o intelecto, dependência ao sujeito. É uma doença hereditária, progressiva, de manifestação tardia (geralmente na idade adulta) e devida a uma mutação em um gene localizado no cromossomo 14q32.1 (7).

O quadro clínico é dominado por falta de coordenação motora, que geralmente inicia-se pela marcha, atingindo a fala, que se torna pouco nítida e, então, atinge os movimentos finos das mãos. Outras manifestações freqüentes são as alterações oculares, que englobam a limitação dos movimentos dos olhos. Com o tempo, a doença se agrava ou se mantém estável apenas. Os pacientes não se movem, têm dificuldade para falar e ficam confinados à cadeira de rodas (8).

A DMJ é também conhecida como doença Açoriana. Essa designação provém do fato de o primeiro caso da doença ter sido verificado em famílias norte-americanas de descendência luso-açoriana nos anos de 1970. Além disso, a Ilha das Flores, nos Açores, é o local de maior incidência da doença em todo mundo.

Uma coorte de pacientes do Hospital de Clínicas de Porto Alegre (HCPA) portadores da DMJ foi acompanhada durante um período de 10 anos para estudar a história natural da doença, através de avaliações clínicas realizadas com esses pacientes. A mensuração do comprometimento clínico do paciente foi efetuada através do escore de exame neurológico para ataxia spinocerebellar (*Neurological Examination Score for Spinocerebellar Ataxia - NESSCA*), que é um escore baseado em uma avaliação quantitativa do exame neurológico padronizado e seu foco está nas principais características, em geral, da SCA e, em particular, da DMJ (9).

As avaliações clínicas dos pacientes portadores da DMJ representam dados longitudinais não-regulares com estrutura desbalanceada. Com base nesses dados há o interesse de avaliar o comportamento da NESSCA ao longo do tempo. Assim, o objetivo deste trabalho é ajustar curvas de crescimento evolução para descrever a progressão da DMJ, quantificada pelo escore NESSCA, utilizando o modelo de coeficientes aleatórios.

MÉTODOS

Os dados utilizados neste trabalho foram provenientes de um estudo procedido no HCPA, entre maio de 1995 e junho de 2005, com uma amostra de 105 sujeitos cujo diagnóstico é positivo para a DMJ. Cada consulta médica foi realizada de acordo com a procura e/ou retorno do sujeito ao atendimento, o que não ocorreu em todos os anos subsequentes à primeira consulta.

A mensuração do comprometimento do sujeito devido à doença foi procedida através da escala NESSCA. Esta escala é composta por 18 itens produzindo um escore total que varia de 0 a 40, onde zero é nenhum comprometimento. Treze desses itens correspondem à parte de exame neurológico padronizado e cinco são subjetivos contando com as informações dos pacientes. Nas avaliações dessa escala, todos os sujeitos foram entrevistados e examinados através da NESSCA pela mesma médica (L.B.J.), caracterizando o procedimento padronizado de avaliação que incluiu um exame neurológico completo e também a determinação de algumas covariáveis, como a idade de início da doença, o comprimento da mutação, entre outros (8;9). Ressalva-se que diante das observações houve, para alguns sujeitos, uma imputação clínica do tipo recordatória. Essa imputação realizou-se para observações faltantes no intervalo entre uma avaliação e outra, quando o escore NESSCA manteve-se constante entre as duas avaliações. E, ainda, no caso da alteração no escore NESSCA não ser considerada brusca, permitindo através dos sintomas relatados pelo

paciente estimar-se “o momento em que houve alteração no escore”.

O comprimento da mutação é definido como seqüência repetitiva CAG do alelo ATXN3 expandido. Todas as pessoas têm nos dois alelos do gene ATXN3 (gene que quando mutado causa a DMJ) uma seqüência repetitiva de codons CAG, que no caso dos portadores da DMJ é uma seqüência repetitiva aumentada, dita, expandida. Sujeitos sem a DMJ têm entre 10 a 40 repetições nesta seqüência, já portadores da DMJ têm entre 62 a 86. Portanto, a mensuração desta seqüência repetitiva aumentada definiu a variável comprimento da mutação. Com relação ao início da doença, considerado como o aparecimento dos primeiros sintomas, ressalta-se que este foi datado pelo relato do próprio sujeito ou familiar próximo.

Quanto à análise estatística dos dados, inicialmente procedeu-se uma análise exploratória com objetivo de verificar o comportamento dos sujeitos para determinadas variáveis. As características de tendência central e de variabilidade, tais como, média, mediana, desvio padrão, máximo, mínimo e outras, foram verificadas para o comprimento da mutação, a idade no início da doença e o escore NESSCA. Também foram efetuados gráficos de linha e dispersão. Para efetuar essas análises iniciais foi utilizado o software SPSS (*Statistical Package for the Social Sciences*).

Posteriormente foi realizado o ajuste ao modelo de coeficientes aleatórios que, por sua vez, descreve aritmeticamente a relação entre as observações e o tempo. Nesse modelo são estimadas as curvas de crescimento para cada sujeito, e também as curvas de crescimento dos grupos, que consideram os parâmetros das curvas de crescimentos individuais, e, ainda, a curva de crescimento média. Neste trabalho, pressupôs-se a relação linear entre o desfecho (escore NESSCA) e o tempo, como acontece nas aplicações mais comuns do modelo. Dessa forma, as curvas de crescimento ajustadas são, na verdade, retas. É importante destacar que a flexibilidade do modelo de coeficientes aleatórios para estrutura de dados desbalanceada foi imprescindível para este trabalho, porque, conforme já dito anteriormente, não aconteceram consultas médicas em todos os anos subsequentes à primeira consulta. Para mais detalhes sobre este modelo, ver Brown e Prescott (6), Bergamo e Corrente (10) e Goldstein (11). O modelo misto, conforme mencionado anteriormente, pode também ser chamado de modelo multinível. Nesta aplicação, isso ocorre porque há dois níveis envolvidos: sujeitos e avaliações. O modelo de coeficientes aleatórios é um caso especial do modelo misto, pois os coeficientes de regressão das curvas individuais serão considerados efeitos aleatórios e os coeficientes de regressão das curvas médias, efeitos fixos.

Assim sendo, com objetivo de avaliar a progressão da DMJ ao longo do tempo, utilizou-se o escore NESSCA como desfecho, e buscou-se verificar se a progressão da doença poderia ser influenciada pelas variáveis explicativas idade no início da doença e comprimento da

mutação. Foram ajustados modelos tendo como variáveis explicativas, ora comprimento da mutação, ora idade no início da doença e, por fim, ambas as variáveis (comprimento da mutação e idade inicial) dicotomizadas. A seguir, a formulação dos três modelos é apresentada.

Modelo 01 - Idade no início da doença contínua

$$nessca_{ij} = (\alpha_0 + \alpha_1 idinic_i + U_{1i}) + (\beta_0 + \beta_1 idinic_i + U_{2i}) tempo_{ij} + \varepsilon_{ij}$$

Onde:

$i = 1, \dots, n$, onde n é o número de sujeitos, neste caso, $n=105$;

$j = 1, \dots, n_i$, onde n_i é o número de observações do sujeito i ;

$nessca_{ij}$ é o escore da escala NESSCA do sujeito i na avaliação j ;

$idinic_i$ é a idade no início da doença do sujeito i ;

U_{1i} e U_{2i} são efeitos aleatórios de cada sujeito, U_{1i} mede o quanto o intercepto da reta do i -ésimo sujeito se afasta do intercepto da reta média, e analogamente, U_{2i} mede este afastamento em termos da inclinação;

$tempo_{ij}$ é o tempo desde o início da doença para o sujeito i na avaliação j ;

ε_{ij} é o erro aleatório.

Para completar a especificação deste modelo pressupõe-se que os efeitos aleatórios tenham distribuição normal bivariada com média zero e matriz de variâncias-covariâncias dita não estruturada:

$$U_i = \begin{bmatrix} U_{1i} \\ U_{2i} \end{bmatrix} \sim N(0, D) \quad D = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \quad \text{e também,} \quad \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

Esta especificação permite modelar a correlação entre avaliações de um mesmo sujeito.

Note que, como $E[U_{1i}] = E[U_{2i}] = E[\varepsilon_{ij}] = 0$, os efeitos fixos são interpretados como coeficientes de regressão da reta média:

$$E[nessca_{ij}] = (\alpha_0 + \alpha_1 idinic_i) + (\beta_0 + \beta_1 idinic_i) tempo_{ij} \quad [1]$$

Onde:

α_0 é o intercepto da reta média dos sujeitos com idade no início da doença igual a zero, o que não é clinicamente relevante;

α_1 é a variação no intercepto da reta média dos sujeitos dada uma unidade de variação na idade no início da doença do sujeito;

β_0 é a inclinação da reta média dos sujeitos com idade no início da doença igual a zero;

β_1 é a variação na inclinação da reta média dos sujeitos dada uma unidade de variação na idade no início da doença do sujeito.

Neste modelo, os efeitos aleatórios permitem que, além das covariáveis, características individuais também sejam consideradas. A variabilidade dos efeitos aleatórios é denominada de variabilidade entre sujeitos, pois mede o quanto sujeitos diferentes têm evoluções da

doença diferentes. Já a variabilidade do erro aleatório é denominada de variabilidade intra-sujeitos, pois mede o quanto as observações de um mesmo sujeito se desviam da reta média. Os efeitos fixos $\alpha_0, \alpha_1, \beta_0$ e β_1 , usualmente são os parâmetros de maior interesse em serem estimados.

Modelo 02 - Comprimento da mutação contínuo

De forma análoga ao modelo com a idade inicial, temos para o comprimento da mutação:

$$E[\text{nessca}_{ij}] = (\alpha_0 + \alpha_1 \text{cagexp}_i) + (\beta_0 + \beta_1 \text{cagexp}_i) \text{tempo}_{ij} \quad [2]$$

Onde:

cagexp_i é o comprimento da mutação do sujeito i ;

α_0 é o intercepto da reta média dos sujeitos com comprimento da mutação igual a zero, o que não é clinicamente possível;

α_1 é a variação no intercepto da reta média dos sujeitos dada uma unidade de variação no comprimento da mutação;

β_0 é a inclinação da reta média dos sujeitos com comprimento da mutação igual a zero;

β_1 é a variação na inclinação da reta média dos sujeitos dada uma unidade de variação no comprimento da mutação.

Modelo 03 - Idade no início da doença e comprimento da mutação dicotomizados

$$\text{nessca}_{ij} = (\alpha_0 + \alpha_1 \text{idinic}_i + \alpha_2 \text{cagexp}_i + U_{1i}) + (\beta_0 + \beta_1 \text{idinic}_i + \beta_2 \text{cagexp}_i + U_{2i}) \text{tempo}_{ij} + \varepsilon_{ij}$$

Onde:

nessca_{ij} , tempo_{ij} , U_{1i} , U_{2i} , ε_{ij} são definidos como anteriormente;

$$\text{idinic}_i = \begin{cases} 0, & \text{idinic} < 34 \\ 1, & \text{idinic} \geq 34 \end{cases} \quad \text{cagexp}_i = \begin{cases} 0, & \text{cagexp} < 74 \\ 1, & \text{cagexp} \geq 74 \end{cases}$$

Como não há evidências sobre um ponto de corte adequado para a idade no início da doença e o comprimento da mutação, optou-se pela utilização da mediana.

Com as mesmas suposições dos modelos anteriores em relação à distribuição dos efeitos aleatórios e do erro aleatório, podemos derivar a reta média, conforme abaixo:

$$E[\text{nessca}_{ij}] = (\alpha_0 + \alpha_1 \text{idinic}_i + \alpha_2 \text{cagexp}_i) + (\beta_0 + \beta_1 \text{idinic}_i + \beta_2 \text{cagexp}_i) \text{tempo}_{ij} \quad [3]$$

Como idinic e cagexp são variáveis binárias, pode-se obter as retas médias para cada grupo:

$$E[\text{nessca}_{ij} \mid \text{cagexp} = 0, \text{idinic} = 0] = \alpha_0 + \beta_0 \text{tempo}_{ij}$$

$$E[\text{nessca}_{ij} \mid \text{cagexp} = 0, \text{idinic} = 1] = (\alpha_0 + \alpha_1) + (\beta_0 + \beta_1) \text{tempo}_{ij}$$

$$E[\text{nessca}_{ij} \mid \text{cagexp} = 1, \text{idinic} = 0] = (\alpha_0 + \alpha_2) + (\beta_0 + \beta_2) \text{tempo}_{ij}$$

$$E[\text{nessca}_{ij} \mid \text{cagexp} = 1, \text{idinic} = 1] = (\alpha_0 + \alpha_1 + \alpha_2) + (\beta_0 + \beta_1 + \beta_2) \text{tempo}_{ij}$$

Assim, a interpretação dos parâmetros é a seguinte:

α_0 é o intercepto da reta média dos sujeitos com idade no início da doença inferior a 34 anos e comprimento da mutação inferior a 74 ($\text{idinic} = \text{cagexp} = 0$);

α_1 é a diferença no intercepto da reta média dos sujeitos com idade no início da doença maior ou igual a 34 anos ($\text{idinic} = 1$) em relação aos sujeitos com idade no início da doença inferior a 34 anos ($\text{idinic} = 0$), para sujeitos na mesma categoria de comprimento da mutação;

α_2 é a diferença no intercepto da reta média dos sujeitos com comprimento da mutação maior ou igual a 74 ($\text{cagexp} = 1$) em relação aos sujeitos com comprimento da mutação inferior a 74 ($\text{cagexp} = 0$), para sujeitos na mesma categoria de idade inicial;

β_0 é a inclinação da reta média dos sujeitos com idade no início da doença inferior a 34 anos e comprimento da mutação inferior a 74 ($\text{idinic} = \text{cagexp} = 0$);

β_1 é a diferença na inclinação da reta média dos sujeitos com idade no início da doença maior ou igual a 34 anos ($\text{idinic} = 1$) em relação aos sujeitos com idade no início da doença inferior a 34 anos ($\text{idinic} = 0$), para sujeitos na mesma categoria de comprimento da mutação;

β_2 é a diferença na inclinação da reta média dos sujeitos com comprimento da mutação maior ou igual a 74 ($\text{cagexp} = 1$) em relação aos sujeitos com comprimento da mutação inferior a 74 ($\text{cagexp} = 0$), para sujeitos na mesma categoria de idade inicial.

Então, diante das equações apresentadas para cada modelo e após a estimação dos parâmetros dos mesmos, tornou-se possível a representação gráfica da reta média ajustada em cada o modelo. Em se tratando do modelo somente para idade no início da doença, optou-se por fixar os valores da variável com base nos quintis e, assim, plotar a reta média para cada quintil permitindo, dessa forma, a visualização de qualquer alteração na reta média com a mudança na idade no início da doença. O mesmo método foi utilizado para o modelo com apenas o comprimento da mutação. Já para o modelo com idade no início da doença e comprimento da mutação dicotomizados, tem-se quatro equações que apontam as categorias das variáveis dicotomizadas, logo, a partir dessas equações foi possível construir um gráfico com a reta média de cada categoria. Os gráficos que expressam as retas médias dos modelos ajustados foram confeccionados no software Excel e serão apresentados na próxima seção.

Para analisar os modelos acima especificados, foi utilizado o procedimento Proc MIXED do software SAS (Statistical Analysis Software). Singer (12) comenta que o Proc MIXED do SAS é um procedimento flexível para o ajuste de modelos com coeficientes aleatórios. Isso pelo fato desse procedimento ter sido desenvolvido para analisar modelos mistos, em que são levados em conta efeitos fixos e aleatórios, permitindo considerar interceptos e inclinações como efeitos aleatórios e medir a variabilidade existente dentro dos sujeitos e entre sujeitos. Para fins de exemplo, segue abaixo a rotina utilizada para o modelo 1 (os termos em negrito referem-se aos nomes das variáveis no banco de dados e o termo em itálico para o nome do arquivo do banco de dados):

```
proc mixed data=work.nessca METHOD=REML CL
ALPHA=.05;
class id;
model nessca = tempo idinic tempo*idinic/ddfm
= satterth solution cl;
random intercept tempo / sub=id type=UN solu-
tion cl;
run;
```

A opção `class id`, onde `id` é a variável de identificação dos sujeitos, indica que os dados representam observações múltiplas ao longo do tempo para sujeitos. É nesta opção que se indica quais são as covariáveis categóricas. `model` indica os efeitos fixos e `random` os efeitos aleatórios. Então, em `model` temos a especificação da variável dependente (**nessca**) e também do **tempo**, **idinic** e a interação **tempo*idinic** que são os efeitos fixos. Para `random` temos os termos `intercept` e **tempo** referentes ao intercepto e a inclinação da reta individual, respectivamente. A opção `sub=id`

permite a variação entre os interceptos dos sujeitos e também entre as inclinações das retas de cada sujeito. Em `type=UN` temos a opção por uma matriz de variâncias-covariâncias de tipo não-estruturada para os efeitos aleatórios.

A especificação `METHOD=REML` considera que o método de estimação dos componentes de variâncias utilizado foi o REML (método da máxima verossimilhança restrita, ou residual), uma vez que esse método produz estimativas não-viciadas (5). O comando `ALPHA=.05` define o nível de confiança de 95%, `solution` solicita a exibição das estimativas dos efeitos e `cl` a exibição do intervalo de confiança. A opção `ddfm = satterth` refere-se ao procedimento empregado para aproximar os graus de liberdade (GL), neste caso, o procedimento utilizado foi o Satterthwaite. Mais informações sobre o ajuste desse modelo e outros modelos para dados longitudinais no SAS podem ser encontrados em Brown (6); Verbeke&Molenberghs (13); Littell (14) e Singer (12).

RESULTADOS

Na tabela 1 encontram-se os valores das estatísticas descritivas para as variáveis estudadas: idade no início da doença (`idinic`), comprimento da mutação (`cagexp`) e escore NESSCA.

Considerando os resultados para a amostra inteira apresentados na Tabela 1, salienta-se a grande amplitude da idade no início da doença, variando de 7 a 57 anos. Ou seja, nota-se a existência de uma grande variabilidade para a idade no início da doença nos sujeitos.

Em se tratando das análises com a amostra dicotomizada pela variável comprimento da mutação (`cagexp`), na Tabela 1 vemos que os sujeitos com comprimento da mutação inferior a 74, possuem, em média, idade no início da doença maior e escore NESSCA menor, em relação aos sujeitos com comprimento da mutação maior ou igual a 74. No caso da amostra dicotomizada pela variável idade no início da doença (`idinic`), percebe-se que os sujeitos com idade no início da doença inferior a 34 anos apresentam, em média, escore NESSCA maior em relação aos sujeitos com idade no início da doença maior ou igual a 34 anos. É importante destacar que nesses resultados o fator tempo de evolução da doença não está sendo considerado, logo há apenas uma sugestão de que uma menor idade no início da doença e um maior comprimento da mutação estejam relacionados com um maior grau de comprometimento na progressão da DMJ. Apenas o modelo longitudinal pode testar essas diferenças de forma fidedigna.

Na Figura 1 temos as curvas para cada sujeito conforme o escore NESSCA observado em cada ano de evolução da doença.

Tabela 1 - Estatísticas descritivas

	Variável	N	Média	Mediana	Mínimo	Máximo	Desvio Padrão
Amostra inteira	cagexp	100	74,24	74	67	82	2,65
	idinic	105	34,09	34	7	57	10,07
	NESSCA	105	16,43	16	3	33	5,20
CAGEXP < 74	cagexp	45	71,89	72	67	73	1,22
	idinic	45	39,00	38	24	57	7,70
	NESSCA	45	15,37	16	3	29	4,98
CAGEXP \geq 74	cagexp	55	76,16	76	74	82	1,84
	idinic	55	29,82	29	15	50	9,17
	NESSCA	55	17,51	17	5	33	5,34
IDINIC < 34	cagexp	48	75,44	76	71	82	2,73
	idinic	49	25,10	26	7	33	5,70
	NESSCA	49	17,33	17	3	33	5,20
IDINIC \geq 34	cagexp	52	73,13	73	67	77	2,03
	idinic	56	41,95	42	34	57	5,40
	NESSCA	56	15,76	16	3	31	5,11

No gráfico apresentado na Figura 1 é possível observar que o escore NESSCA possui comportamento crescente ao longo dos anos de evolução da doença. Além disso, verifica-se que as curvas individuais apresentam perfis bastante

diferentes, mostrando que as manifestações da doença variam de sujeito para sujeito. Ou seja, indicando, assim, que há grande variabilidade entre sujeitos.

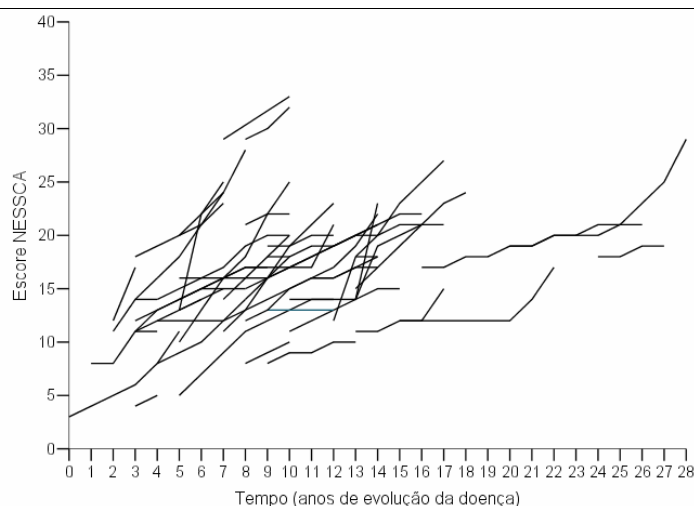


Figura 1 - Gráfico das curvas individuais para NESSCA.

Na Figura 2 temos, as curvas do escore NESSCA médio em cada ano, separadas pela categoria de idade de início da doença. Na Figu

ra 3 tem-se um gráfico similar, onde as curvas estão separadas pelas categorias de comprimento da mutação.

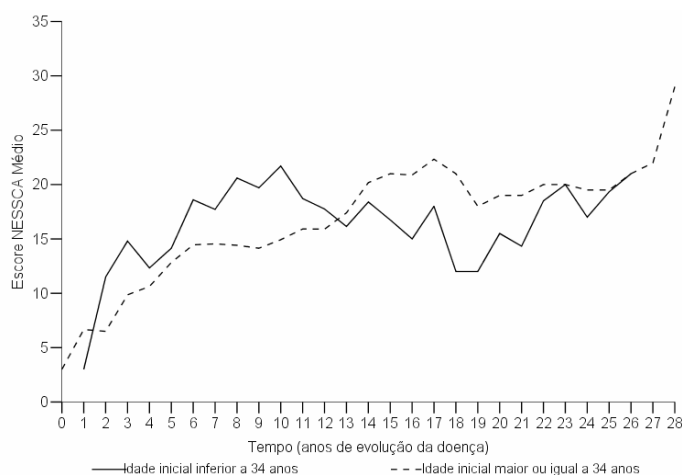


Figura 2 - Gráfico das curvas médias para idade no início da doença dicotomizada.

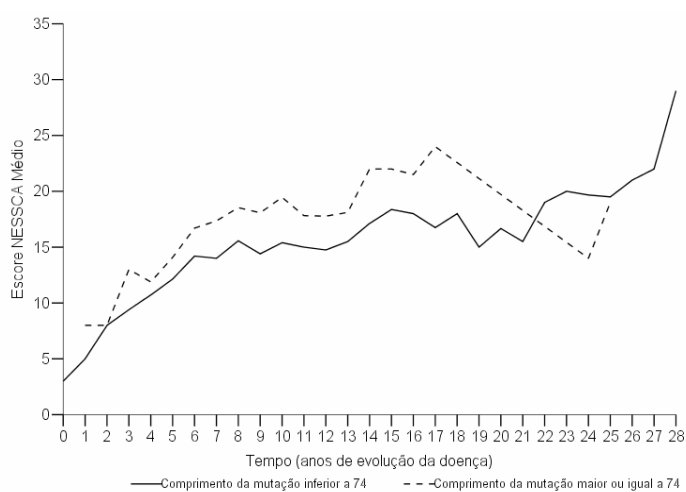


Figura 3 - Gráfico das curvas médias para comprimento da mutação dicotomizada.

Nas Figuras 2 e 3 nota-se que a NESSCA média, para ambas as classes é em geral crescente ao longo dos anos de evolução da doença. Além disso, na Figura 3, a curva referente aos sujeitos com comprimento da mutação inferior a 74 inicia antes e termina depois da curva dos sujeitos com comprimento maior ou igual a 74. Contudo, a exceção dos tempos superiores a 21, esta última é superior a curva dos sujeitos com comprimento inferior a 74 em praticamente todos os tempos, sugerindo, deste modo, que

um maior comprimento da mutação leva a uma progressão da doença mais grave e que temos uma observação discrepante das demais no tempo 24. Já na Figura 2 esta distinção entre as classes não está clara, visto que os perfis médios se cruzam.

Considerando a variável idade no início da doença de forma contínua, idinic, e seguindo o modelo 1, as estimativas dos parâmetros considerados são mostradas na Tabela 2.

Tabela 2 – Estimativa dos parâmetros para o ajuste da NESSCA à idade de início.

Efeito	Parâmetro	Estimativa	Erro Padrão	valor-p	IC 95%
Intercepto	α_0	8,34	2,89	0,01	(2,56 ; 14,12)
tempo	β_0	1,87	0,34	<0,01	(1,19 ; 2,55)
idinic	α_1	-0,06	0,08	0,45	(-0,23 ; 0,10)
tempo*idinic	β_1	-0,02	0,01	0,03	(-0,04 ; -0,00)

A reta média estimada tem então a seguinte forma:

$$\hat{E}[\text{nessca}_{ij}] = (8,34 - 0,06\text{idinic}_i) + (1,87 - 0,02\text{idinic}_i)\text{tempo}_{ij}$$

De acordo com a Tabela 2, estima-se que, com um ano a mais na idade inicial, a NESSCA observada no início da doença diminua 0,06 unidades, mas esse efeito não é significativo a 5%. Entretanto, temos diferença significativa observada na taxa de crescimento ao longo do tempo, ou seja, na inclinação da reta média. Assim, estima-se que com o acréscimo de uma unidade na idade no início da doença do sujeito, ocorre a diminuição de 0,02 na inclinação da reta média. Desta forma, é possível afirmar que quanto maior a idade inicial, menor é o acréscimo na inclinação da reta, portanto, de forma mais lenta é a progressão da doença.

A matriz de variâncias-covariâncias dos efeitos aleatórios estimada é dada por:

$$D = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 35,75 & -2,88 \\ -2,88 & 0,47 \end{bmatrix}$$

A variação nos interceptos é 35,75, enquanto que 0,47 representa a variação nas inclinações. Isso indica que a variabilidade entre os sujeitos é maior no que diz respeito ao escore da NESSCA no início da doença do que na velocidade de evolução da doença. Já a covariância entre os interceptos e as inclinações é -2,88, indicando relação negativa entre eles. Além disso, a variância residual (variabilidade intra-sujeitos) é igual a 0,95.

Os resultados do modelo para a idade no início da doença contínua são apresentados graficamente na Figura 4.

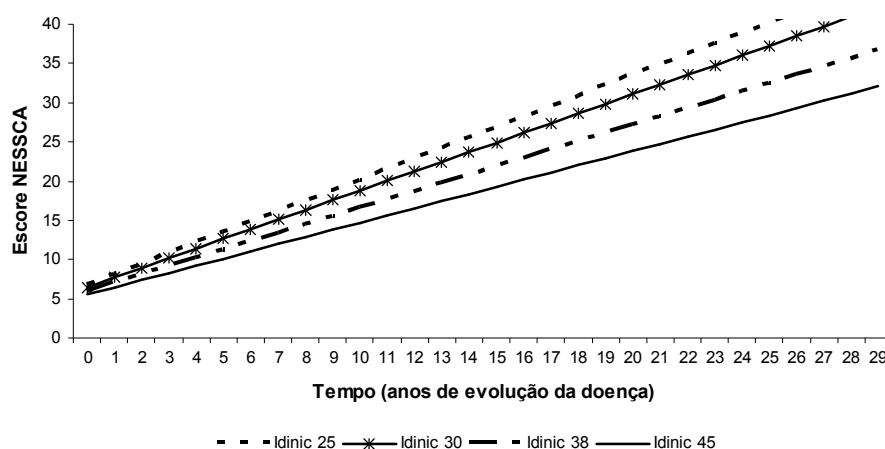


Figura 4 - Gráfico das retas ajustadas para a idade no início da doença (Quintis).

Com base na Figura 4 que representa as retas médias para os quintis observados de idade no início da doença, observa-se que, com o aumento da idade inicial, menores são os interceptos e também as inclinações das retas, ou seja, quanto mais tarde a doença se manifesta, menor é o valor do escore NESSCA no início da doença e a progressão é mais lenta. Estes

aspectos sugerem que a manifestação tardia dos sintomas da doença pode ser fator protetor para a doença.

Considerando a variável comprimento da mutação de forma contínua, *cagexp*, e seguindo o modelo 2, as estimativas dos parâmetros observados são mostradas na Tabela 3.

Tabela 3 – Estimativa dos parâmetros do ajuste da NESSCA ao comprimento da mutação.

Efeito	Parâmetro	Estimativa	Erro Padrão	valor-p	IC 95%
Intercepto	α_0	37,99	23,56	0,11	(-9,17 ; 85,15)
tempo	β_0	-11,12	2,43	<0,01	(-15,96 ; -6,28)
<i>cagexp</i>	α_1	-0,43	0,32	0,18	(-1,06 ; 0,21)
tempo* <i>cagexp</i>	β_1	0,16	0,03	<0,01	(0,10 ; 0,23)

Temos então que a reta média estimada é dada por:

$$\hat{E}[\text{nessca}_{ij}] = (37,99 - 0,43\text{cagexp}_i) + (-11,12 + 0,16\text{cagexp}_i)\text{tempo}_{ij}$$

Note que α_0 e β_0 não são plausíveis porque representam, respectivamente, o intercepto e a inclinação da reta média dos sujeitos com comprimento da mutação igual a zero. Entretanto, sabe-se que em portadores da DMJ o comprimento da mutação é no mínimo 62.

Os resultados da Tabela 3 mostram que, considerando os coeficientes para cagexp, apenas o coeficiente relacionado à variação na inclinação da reta média para cagexp é significativo a 5%. Assim, estima-se que, com o acréscimo de uma unidade no comprimento da mutação, ocorre o aumento de 0,16 na inclinação da reta média. Ou seja, quanto maior o comprimento da mutação mais rápida acontece a progressão da doença.

A matriz de variâncias-covariâncias dos efeitos aleatórios estimada é:

$$D = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 35,62 & -2,12 \\ -2,12 & 0,29 \end{bmatrix}$$

Assim como no modelo anterior, observa-se relação negativa entre os interceptos e inclinações das retas individuais. Como o valor da variância residual foi 0,98 também temos novamente que há mais variabilidade entre sujeitos que dentro de sujeitos.

O modelo ajustado para o comprimento da mutação contínuo é apresentado graficamente pela Figura 5.

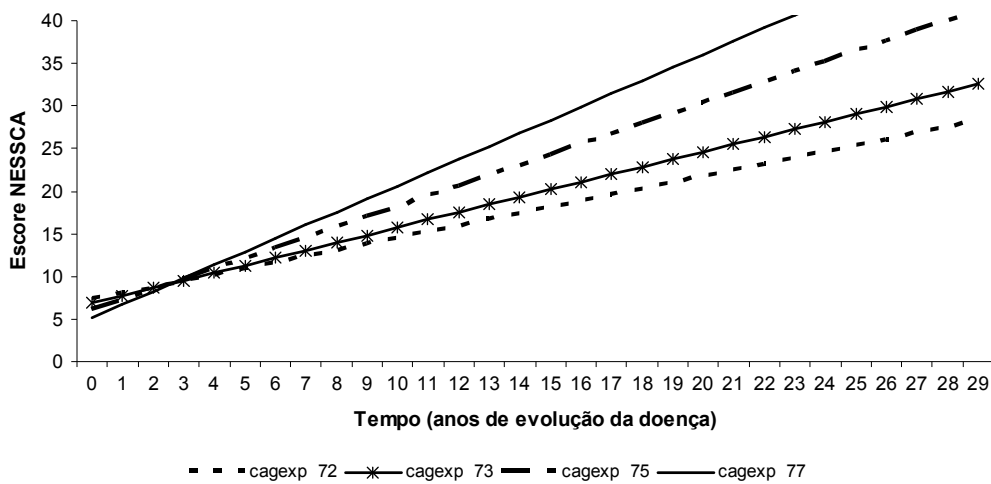


Figura 5 - Gráfico das retas ajustadas para o comprimento da mutação (Quintis).

O gráfico das retas ajustadas para o comprimento da mutação apresentado pela Figura 5, mostra características peculiares, uma vez que, com o aumento do comprimento há diminuição no valor do intercepto e elevação na inclinação. Desta forma, sujeitos com comprimento da mutação alto possuem valor de escore NESSCA baixo no início da doença e progressão rápida. Por outro lado, os sujeitos com comprimento da mutação baixo, no início da doença, apresentam valores do escore NESSCA mais altos e progredem de forma mais lenta. Esses resultados sugerem que o comprimento da mutação, quando referido a inclinação pode ser considerado fator de risco para a doença.

O fato de o modelo estimar que o aumento de uma unidade no comprimento da mutação faz diminuir o intercepto da reta média, apesar de estatisticamente não significativo, contraria o esperado pela prática médica. Acredita-se, então, que este resultado seja devido a um viés recordatório, já que o início da doença foi determinado através dos próprios sujeitos. Há indícios de que sujeitos com comprimento da mutação mais altos adoecem mais precocemen-

te, na juventude ou mesmo na infância. Neste estudo, temos que a correlação linear de Pearson entre idade no início da doença e comprimento da mutação foi de -0,65 ($p < 0,001$), indicando relação negativa entre ambas as variáveis. Então, esses sujeitos adoecem em um período no qual são observados e cuidados por seus pais (entre eles, um doente, visto que a doença é hereditária) e familiares. Portanto, é possível que isso os leve mais cedo ao atendimento médico e, conseqüentemente, que as informações dadas por eles sobre o início dos sintomas sejam mais precisas. Já para sujeitos que procuram atendimento mais tarde, é mais difícil lembrar a exata data do início dos sintomas assim, provavelmente, eles lembram apenas de datas dos sintomas iniciais mais graves.

Uma outra possível explicação para o resultado encontrado seriam as diferentes constituições do escore NESSCA para cada categoria de sujeitos. Naqueles com comprimento da mutação maior ou igual a 74, o quadro clínico envolve achados das vias corticospinais e extrapiramidais. Já nos sujeitos com comprimento da mutação inferior a 74, o quadro clínico inclui

manifestações do sistema nervoso periférico. Assim, não é impossível que o escore NESSCA deste último grupo de sujeitos some mais achados do que o do primeiro grupo, na primeira avaliação clínica.

Para as variáveis comprimento da mutação, cagexpd, e idade no início da doença, idinid, dicotomizadas, as estimativas dos parâmetros consideradas são mostradas na Tabela 4, seguindo o modelo 3.

Tabela 4. Estimativa dos parâmetros do ajuste da NESSCA à idade de início e comprimento da mutação dicotomizados.

Efeito	Parâmetro	Estimativa	Erro Padrão	valor-p	IC 95%
Intercepto	α_0	7,85	1,84	<0,01	(4,16 ; 11,54)
tempo	β_0	0,88	0,19	<0,01	(0,50 ; 1,25)
idinid	α_1	-0,73	1,92	0,71	(-4,57 ; 3,12)
cagexpd	α_2	-2,44	1,92	0,21	(-6,29 ; 1,41)
tempo*idinid	β_1	-0,15	0,19	0,44	(-0,54 ; 0,24)
tempo*cagexpd	β_2	0,59	0,19	<0,01	(0,21 ; 0,98)

A reta média ajustada pode ser expressa da seguinte forma:

$$\hat{E}[\text{nessca}_{ij}] = (7,85 - 0,73\text{idinid}_i - 2,44\text{cagexpd}_i) + (0,88 - 0,15\text{idinid}_i + 0,59\text{cagexpd}_i)\text{tempo}_{ij}$$

A Tabela 4 informa que as estimativas para o intercepto e também para a inclinação, quando a idade no início da doença é inferior a 34 anos e o comprimento da mutação é inferior a 74, são significativamente diferentes de zero ao nível de significância 5%. Ou seja, para este grupo, demonstra-se a relação crescente da NESSCA com o tempo. Outra estimativa que se mostra significativa a 5% refere-se à variação no coeficiente de inclinação para o comprimento da mutação, quando o mesmo é maior ou igual a 74. Então, é possível afirmar, considerando uma mesma categoria de idade inicial, que os sujeitos com comprimento da mutação maior ou igual a 74 possuem inclinação 0,59 unidades superior, em relação aos sujeitos de comprimento da mutação inferior a 74. Portanto, as informações sugerem que os sujeitos com comprimento da mutação inferior a 74 possuem progressão da doença de forma mais lenta em relação a outra classe.

No que diz respeito a idade inicial, nem a variação no intercepto nem na inclinação, para a idade no início da doença maior ou igual a 34 anos, são significativos.

A matriz de variâncias-covariâncias dos efeitos aleatórios estimada é apresentada da seguinte forma:

$$D = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 37,65 & -2,53 \\ -2,53 & 0,41 \end{bmatrix}$$

Observa-se para este modelo, as mesmas conclusões dos modelos anteriores, sendo, neste caso, a variância residual é igual a 0,94.

A Figura 6 mostra de forma gráfica os resultados obtidos para a idade no início da doença e também o comprimento da mutação dicotomizados.

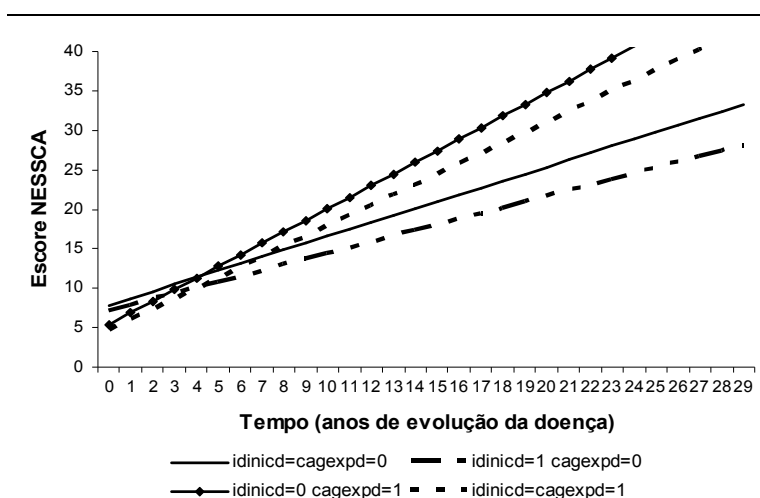


Figura 6 - Gráfico das retas ajustadas para idade no início da doença e comprimento da mutação dicotomizados.

Conforme a Figura 6, observa-se que sujeitos com comprimento da mutação maior ou igual a 74, independentemente da categoria referente a idade do início da doença, possuem os menores interceptos e maiores inclinações. Ou seja, são os sujeitos que iniciam com os escores NESSCA mais baixos, mas possuem a progressão da doença ao longo dos anos de forma mais rápida. Em contrapartida, aqueles com comprimento da mutação inferior a 74, também independente da categoria da idade do início da doença, possuem os escores da NESSCA mais altos no início da doença, mas a progressão da doença mais lenta. Esse resultado sobre as inclinações vai ao encontro das hipóteses clínicas, mas a questão dos interceptos não, e novamente pensa-se que provavelmente isso é devido ao viés de recordação.

Com relação a idade inicial, verifica-se que os sujeitos com idade no início da doença inferior a 34 anos iniciam a doença com um escore NESSCA ligeiramente mais alto e possuem a progressão da doença de forma mais rápida em relação aos sujeitos com idade no início da doença maior ou igual a 34 anos, independente da categoria de comprimento da mutação.

DISCUSSÃO

Neste trabalho descreveu-se a aplicação de um modelo misto no ajuste de curvas de crescimento. Além da descrição geral do modelo aplicado foram apresentadas três alternativas para o ajuste das curvas de crescimento conforme as variáveis indicadas para a modelagem do escore NESSCA dos pacientes portadores da DMJ. Dessa forma, mostramos e detalhamos o modelo e também os métodos utilizados, buscando contribuir, assim, para o aprendizado e compreensão por parte de outros pesquisadores.

O modelo misto mostrou-se eficiente no ajuste das curvas de crescimento para as três alternativas apresentadas. Por esse motivo, é possível concluir que as dificuldades relacionadas ao desbalanceamento dos dados foram superadas.

Os resultados mostraram que quanto mais tardio o início da doença, mais lenta ocorre a sua progressão, sugerindo, assim, que a idade no início da DMJ pode ser considerada como fator de proteção. No caso do comprimento da mutação, quanto maior é este comprimento, mais rápida é a progressão da doença, portanto, há evidências de que o comprimento da mutação pode ser apontado como fator de risco para a DMJ. As atribuições quanto a fator protetor e fator de risco são consideradas somente para a progressão da doença, onde o modelo apontou diferenças significativas. Para o escore NESSCA inicial não há qualquer diferença significativa

encontrada nos ajustes realizados, mas existe a possibilidade de que alguns desses resultados foram afetados por viés de recordação. Quando na observação do ajuste para ambas as variáveis de forma dicotomizada, a idade no início da doença deixou de ter efeito significativo, entretanto é um fator clinicamente relevante e de fácil acesso para predizer a progressão de um paciente. Já o comprimento da mutação mostrou-se impositivo à idade inicial, pois mostrou efeito significativo mesmo na presença da idade inicial.

Como uma continuidade deste estudo, seria interessante utilizar um modelo que não assumisse linearidade entre NESSCA e o tempo, para testar a hipótese de que um portador da DMJ tem uma velocidade maior de progressão nos primeiros anos da doença e depois uma estabilização. Então, ao invés da utilização de modelos mistos lineares trabalhar-se-ia com modelos mistos não lineares.

REFERÊNCIAS

1. Suyama E. Modelos de Efeitos Aleatórios para Dados Longitudinais [Reunião da RBRAS e 6º SEAGRO] Ribeirão Preto, São Paulo; 1995.
2. Scheiner SM, Gurevitch J. Design and Analysis of Ecological Experiments. Second ed. Oxford University Press, Inc; 2001.
3. Burchinal M, Nelson L, Poe M. Growth Curve Analysis: An introduction to various methods for analyzing longitudinal data. Monographs of the Society for Research in Child Development 2006;65-87.
4. Riboldi J. Modelos Mistos Aplicados [52ª RBRAS e 12ª SEAGRO] Santa Maria, Rio Grande do Sul; 2007.
5. Leotti VB, Ziegelmann PK, Riboldi J, Vigo A. Comparação via Simulação dos Estimadores Clássicos e Bayesianos no Modelo de Coeficientes Aleatórios para Dados Longitudinais [Dissertação para obtenção do título de Mestre] Programa de Pós-graduação em Epidemiologia, UFRGS; 2007.
6. Brown H, Prescott R. Applied Mixed Models in Medicine. Second ed. Chichester: John Wiley & Sons, Ltda; 2006.
7. Kieling C, Prestes PR, Saraiva-Pereira ML, Jardim LB. Survival estimates for patients with Machado-Joseph disease (SCA3). Clinical Genetics 2007;72:543-5.
8. Jardim LB, Pereira ML, Silveira I, Ferro A, Sequeiros J, Giugliani R. Neurologic findings in Machado-Joseph disease - relation with disease duration, subtypes and (CAG)n. Archives of neurology 2001;58:899-904.
9. Kieling C, Rieder CRM, Silva ACF, Saute JAM, Cecchin CR, Monte TL, et al. A neurologic examination score for the assessment of spinocerebellar ataxia 3 (SCA3). European Journal of Neurology 2008;15:371-6.

10. Bergamo GC, Corrente JE. Aplicação de Modelos Multiníveis na Análise de Dados de Medidas Repetidas no Tempo [Dissertação para a obtenção do título de Mestre em Agronomia] Escola Superior de Agricultura "Luiz Queiroz", Universidade de São Paulo; 2002.
11. Goldstein H. Multilevel Statistical Models. London: Arnold; 1999.
12. Singer JD. Using SAS proc mixed to fit multilevel models, hierarchical models, and individual growth models. Journal of Educational and Behavioral Statistics 1998;24:323-55.
13. Verbeke G, Geert M. Linear Mixed Models for Longitudinal Data. Kessel Lo: Springer; 2000.
14. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schanbenberger O. SAS for Mixed Models. Second ed. Cary: SAS Institute Inc; 2006.

Recebido: 14/03/2009

Aceito: 20/04/2009