

Marcelo Militão Abrantes

Diferentes estratégias de modelagem de dados longitudinais desbalanceados

Belo Horizonte – MG
Dezembro 2007

Marcelo Militão Abrantes

Diferentes estratégias de modelagem de dados longitudinais desbalanceados

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Mestre em Estatística

Orientador: Prof. Enrico Antônio Colosimo
Co-orientadores: Prof. Marcel de Toledo Vieira
Prof. Eduardo Araújo de Oliveira

MESTRADO EM ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA
INSTITUTO DE CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DE MINAS GERAIS

Belo Horizonte – MG
Dezembro 2007

AGRADECIMENTOS

Prof. Enrico Colosimo, meu orientador, pelo exemplo de conhecimento teórico e aplicado, disponibilidade e paciência.

Prof. Marcel Vieira, meu co-orientador, pelos ensinamentos e disponibilidade para reuniões (mesmo em Belo Horizonte).

Prof. Eduardo Araújo de Oliveira, pela disponibilidade, cessão do banco de dados, ajuda e vontade de exercer uma medicina “realmente” baseada em EVIDÊNCIAS.

Prof. Joel Alves Lamounier pela amizade e exemplo pessoal e profissional que se fazem presentes desde a iniciação científica.

Marta Alice (Martinha), minha esposa, sempre compreensiva e companheira, por fazer parte da minha vida há pouco mais de cinco anos, tornando-a mais fácil, alegre e feliz. Por aceitar me dividir com o computador.

Fulgêncio e Graça, meus pais, por tudo.

Amigos do ICEX que ajudaram nesta dissertação e durante os estudos para as provas: Renata Mattos, Fábio e Helinton

Famílias Gazzinelli, Abrantes e Militão, por serem o berço de tudo.

Gáida e Vô Pedro, pela lembrança sempre presente...

Amigos do Anchieta, onde vivi minha infância, adolescência e ainda vou viver por um bom tempo.

Amigos da Faculdade de Medicina, em especial Fabiano Amaral Fulgêncio da Cunha (Fabinho), Juliano Ferreira Faria (Aza) e Cristiano Menezes Diniz (Cover), com quem dividi os anos de iniciação científica.

SUMÁRIO

| | |
|--|-----------|
| 1 INTRODUÇÃO | 1 |
| 2 MODELOS PARA DADOS LONGITUDINAIS..... | 3 |
| 2.1 Modelo marginal | 3 |
| 2.2 Modelo linear de efeitos mistos (efeitos aleatórios)..... | 4 |
| 2.3 Inferência para os modelos | 8 |
| 2.4 Modelo para dados desbalanceados..... | 8 |
| 2.5 Interpretação dos parâmetros | 9 |
| 3 MOTIVAÇÃO CLÍNICA E MÉTODOS..... | 10 |
| 3.1 Hidronefrose fetal | 10 |
| 3.2 Coleta de dados..... | 12 |
| 3.3 Critérios de inclusão | 12 |
| 3.4 Critérios de exclusão | 12 |
| 3.5 Variáveis consideradas no presente estudo | 13 |
| 3.6 Análise estatística | 14 |
| 4 ESTRATÉGIAS NA ANÁLISE DE DADOS LONGITUDINAIS DESBALANCEADOS..... | 15 |
| 4.1 Introdução..... | 15 |
| 4.2 Dados balanceados | 15 |
| 4.3 Dados desbalanceados | 16 |
| 5 RESULTADOS | 18 |
| 5.1 Descrição da amostra..... | 18 |
| 5.2 Estratégia 1 (volume e covariáveis)..... | 19 |
| 5.3 Estratégia 2 (variável dependente sem delta e covariáveis com delta)..... | 24 |
| 5.4 Comparação da estratégia 1 com estratégia 2..... | 27 |
| 5.5 Estratégia 3 (delta na variável resposta) | 27 |
| 5.6 Estratégia 4 (delta na variável dependente e nas covariáveis)..... | 29 |
| 5.7 Comparação da estratégia 3 com estratégia 4..... | 31 |
| 6 CONCLUSÕES..... | 32 |
| 7 REFERÊNCIAS BIBLIOGRÁFICAS | 33 |
| ANEXOS | 35 |
| ANEXO 1 -Transferir dados para STATA..... | 35 |
| ANEXO 2 – Comandos do STATA | 36 |
| ANEXO 3 – Transformado bancos de dados “largos” em “longos” (uso do SPSS).. | 37 |

LISTA DE ABREVIATURAS

| Abreviatura | Significado |
|-------------|--------------------------------|
| DAP | Diâmetro antero-posterior |
| Dp | Desvio-padrão |
| ITU | Infecção do trato urinário |
| RAP | Medida antero-posterior do rim |
| RL | Medida longitudinal do rim |
| RT | Medida transversal do rim |
| SC | Superfície corporal |

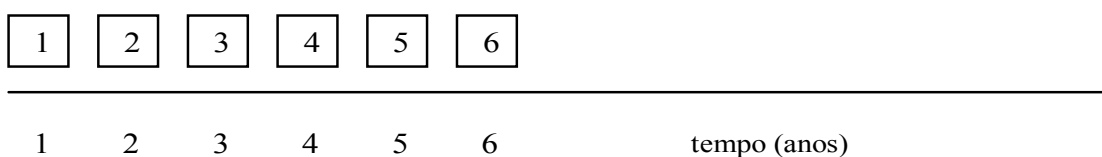
1 INTRODUÇÃO

Estudos longitudinais são aqueles nos quais a variável dependente é medida repetidamente, ou seja, em diferentes ocasiões (mínimo de duas medidas) para o mesmo indivíduo. Em estudos longitudinais as observações de um indivíduo ao longo do tempo usualmente não são independentes entre si, e por isso é necessário uso de técnicas estatísticas especiais que considerem o fato de que medidas repetidas do mesmo indivíduo são correlacionadas (TWISK, 2003).

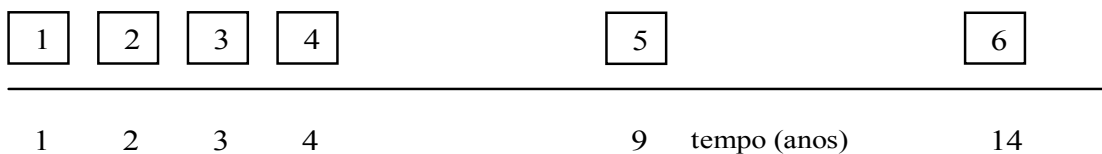
É importante lembrar que estudos de sobrevivência também envolvem uma variável resposta e a observação do tempo. Porém nestes estudos a variável é um evento irreversível e por isso é medida uma única ocasião e após a ocorrência deste evento nenhuma outra observação é feita no indivíduo ou pelo menos considerada na análise dos dados. Assim, a principal vantagem de um estudo longitudinal é permitir a análise de uma determinada variável dependente ao longo do tempo, em termos individuais. Adicionalmente pode-se verificar a influência de outras variáveis (independentes) no comportamento desta variável dependente (TWISK, 2003). Para uma comparação entre dados longitudinais e análises de sobrevivência consultar LIMA (2007).

Dados longitudinais são considerados balanceados quando o número de medidas e o tempo entre estas medidas são os mesmos para cada indivíduo. Por outro lado, quando o número de medidas é diferente entre os indivíduos ou o tempo em que elas foram medidas é diferente os dados são denominados desbalanceados (FIGURA 1).

Medidas em intervalos de tempos iguais



Medidas em intervalos de tempos distintos



Distinção entre dados balanceados e dados desbalanceados

FIGURA 1

Na prática, dados longitudinais são frequentemente desbalanceados no sentido de que um número diferente de medidas está disponível para todos os sujeitos e/ou que as medidas não são obtidas nos mesmos momentos (VERBEKE, 2000).

O objetivo deste trabalho é apresentar diferentes estratégias de modelagem para dados longitudinais desbalanceados. Na Seção 2 vamos apresentar brevemente os dois principais modelos considerados para dados longitudinais e métodos inferenciais. O banco de dados de nefrologia pediátrica que motivou este trabalho é apresentado na Seção 3. As diversas estratégias de modelagem em dados longitudinais avaliadas no presente estudo são apresentadas na Seção 4. Esta dissertação termina na Seção 5 com considerações finais e proposta de trabalhos futuros.

2 MODELOS PARA DADOS LONGITUDINAIS

Os dois modelos clássicos para dados longitudinais, marginal e de efeitos aleatórios, são apresentados nesta seção.

2.1 Modelo marginal

A principal característica deste modelo é que a média e variância da variável aleatória resposta são modeladas separadamente. A interpretação dos coeficientes do modelo é exatamente a mesma daqueles para dados transversais (para cada mudança de uma unidade na variável independente a variável dependente varia em β). A estimação da média será abordada na Seção 2.3. A equação geral do modelo é apresentada abaixo:

$$Y_i = X_i \beta + \varepsilon_i \quad i = 1, \dots, n \quad (1)$$

em que, Y_i representa o vetor resposta de dimensão m_i , do i -ésimo indivíduo, X_i representa a matriz $m_i \times p$ de variável(is) independente(s) de um determinado indivíduo nas observações $j = 1, \dots, m_i$, β representa o p -vetor de parâmetros a serem estimados e também o valor que mudará na variável dependente quando a variável independente apresentar variação de uma unidade, controlada pelas demais variáveis do modelo e ε_i o vetor de erros de dimensão m_i .

Os erros de ajuste do modelo, representados por ε_i apresentam uma distribuição normal m -variada com média zero e variância V_i , ou seja, $\varepsilon_i \sim N(0, V_i)$. A matriz de variância e covariância (V_i) pode ser estruturada ou não estruturada. As principais formas de estruturá-la são apresentadas a seguir:

a) não estruturada: nesta forma são consideradas a variância e covariância para cada par de medidas. Neste caso $\sigma_{kj} = \sigma_{jk}$ e temos $m_i(m_i+1)/2$ componentes de variância

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m_i} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m_i} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m_i1} & \sigma_{m_i2} & \cdots & \sigma_{m_i m_i} \end{bmatrix}$$

$$\sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

b) correlação uniforme: considera-se que a variância é σ^2 e que a covariância entre as medidas é a mesma σ^2 em todos os pares de tempos;

c) auto-regressiva ou correlação serial: semelhante à uniforme, porém a matriz é resultado da multiplicação da variância por um valor σ^2 para as variâncias e por um número representado pelo valor de um coeficiente de correlação elevado à diferença de tempo entre as medidas $\rightarrow (\rho^{|j-k|})$. A intenção é que as informações mais próximas tendem a ser mais correlacionadas que as informações mais distantes pois à medida que aumenta a diferença entre os tempos a correlação diminui. A vantagem é que a matriz é modelada com apenas dois parâmetros (σ^2 e ρ) e $0 < \rho < 1$.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{m_i-1} \\ & 1 & \rho & \dots & \rho^{m_i-2} \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$$

Uma limitação desta modelagem é que não são incorporados fatores individuais que afetem a resposta e possam causar correlação no tempo. Quando se introduz um efeito latente para cada indivíduo pode-se medir quanto o intercepto de um indivíduo difere do restante. Pode-se assumir então que estes efeitos sejam fixos ou aleatórios. O modelo de efeitos fixos assume que os fatores individuais são fixos.

Estas formas de estruturar a matriz de variância e covariância somente podem ser utilizados em dados longitudinais balanceados.

2.2 Modelo linear de efeitos mistos (efeitos aleatórios)

O modelo marginal, apresentado na Seção 2.1, não considera a diferença entre os indivíduos, o que ocorre no modelo de efeitos aleatórios. Esses modelos têm dois componentes: um intraindividual (uma mudança longitudinal intraindividual 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. Neste modelo, as medidas dos indivíduos não precisam ser igualmente espaçadas e balanceadas e as análises podem

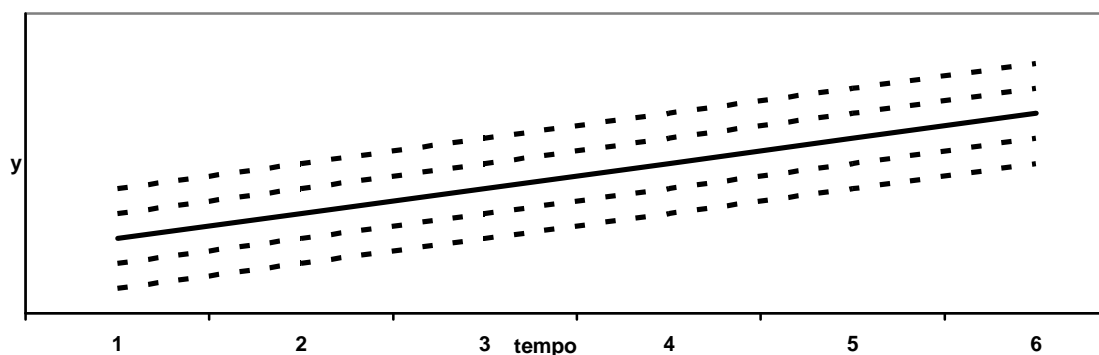
ser conduzidas com os dados de indivíduos que foram perdidos de seguimento ou que apresentam ausência de informação em algum momento do estudo (FAUSTO, 2005).

Será apresentada a seguir a interpretação do modelo com intercepto aleatório somente, do modelo com inclinação aleatória somente e do modelo com intercepto e inclinação aleatórios, respectivamente. Como neste modelo é possível trabalhar com dados desbalanceados o índice do tempo vai variar de 1 até m_i .

No primeiro caso, a forma mais simples de análise de coeficientes aleatórios, considera-se apenas um efeito aleatório e a covariável tempo como única variável independente:

$$\rightarrow Y_{ij} = \beta_{0i} + \beta_1 t_{ij} + \varepsilon_{ij} \quad i = 1, \dots, n \quad \text{e} \quad j = 1, \dots, m_i \quad (2)$$

em que Y_{ij} é a resposta observada para o i -ésimo indivíduo na observação j , β_{0i} é o intercepto aleatório, t_{ij} é o tempo, β_1 é o coeficiente de regressão para o tempo, e ε_{ij} é o “erro” para o i -ésimo indivíduo na observação j . A diferença desta equação em relação a (1) é o intercepto aleatório que indica que o intercepto pode variar entre indivíduos (FIGURA 2) mas mantém-se a inclinação fixa e igual a β_1 .

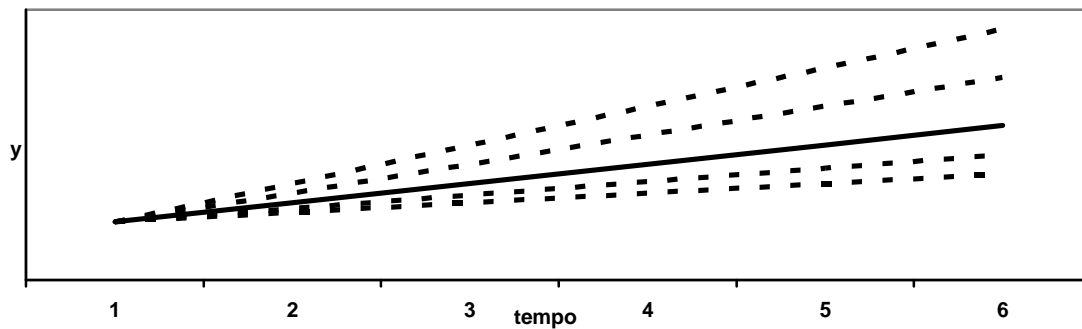


Evolução de uma determinada variável ao longo da resposta (eixo Y); **interceptos diferentes** para indivíduos diferentes (linha tracejada = indivíduos, linha contínua = população)
FIGURA 2

Outra opção é considerar que o intercepto não seja aleatório mas sim que a relação entre a variável aleatória e determinada variável independente seja diferente para cada um dos indivíduos. Ou seja, a inclinação da curva ao longo do tempo seria considerada aleatória, como apresentado na próxima equação:

$$\rightarrow Y_{ij} = \beta_0 + \beta_{1i} t_{ij} + \varepsilon_{ij} \quad i = 1, \dots, n \quad \text{e} \quad j = 1, \dots, m_i \quad (3)$$

em que β_0 é o intercepto fixo, β_{1i} é o coeficiente de regressão aleatório para tempo, e $\rightarrow \varepsilon_{ij}$ é o “erro” para o i -ésimo indivíduo na observação j (FIGURA 3).



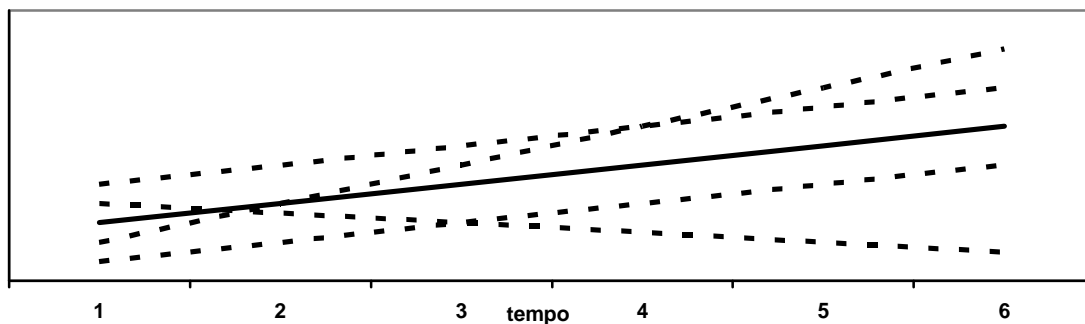
Evolução de uma determinada variável ao longo da resposta (eixo Y); **inclinações diferentes** para indivíduos diferentes (linha tracejada = indivíduos, linha contínua = população)

FIGURA 3

Uma forma mais geral considera a combinação de intercepto e inclinações diferentes ao longo do tempo que é apresentado na equação:

$$\rightarrow Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \varepsilon_{ij} \quad i = 1, \dots, n; \quad j = 1, \dots, m_i \quad (4)$$

em que β_{0i} é o intercepto aleatório, β_{1i} é o coeficiente de regressão aleatório para tempo (FIGURA 4).



Evolução de uma determinada variável ao longo da resposta (eixo Y); **inclinações e interceptos diferentes** para indivíduos diferentes (linha tracejada = indivíduos, linha contínua = população)

FIGURA 4

A forma geral do modelo de efeitos aleatórios (LAIRD e WARE, 1982) é:

$$Y_i = X_i \beta + Z_i b_i + \varepsilon_i, \quad i = 1, \dots, n \quad (5)$$

em que, $X_i \beta$ representa o **componente fixo**; X_i são as covariáveis incluídas na análise e β 's seus respectivos coeficientes de regressão, $Z_i b_i$ representa o **componente aleatório**; b_i 's são os coeficientes aleatórios que variam para cada indivíduo; e, Z_i é, em geral, um subconjunto de X_i , as covariáveis dependentes do tempo e $\varepsilon_i \sim N(0, \sigma^2 I_n)$.

No modelo de efeitos aleatórios assumimos que as respostas do i -ésimo indivíduo têm distribuição Normal m_i -variada com média $(X_i \beta)$ e matriz de covariância $\Sigma = (Z_i V_i Z_i^T + \sigma^2 I)$ em que V_i é a matriz de covariâncias das medidas do mesmo indivíduo. Os coeficientes aleatórios, b_i , são independentes de ε_i e têm distribuição Normal com média zero e matriz de covariância V_i .

Outro comentário importante é que ao se adicionar, por exemplo, o efeito aleatório na inclinação torna-se necessário a estimação de outros dois parâmetros: a variância das inclinações e a covariância entre intercepto aleatório e inclinação aleatória (TWISK, 2003, p. 85). Assim, quando se considera um modelo com efeitos aleatórios no intercepto e na inclinação o modelo tem quatro componentes de variância.

Os efeitos aleatórios no intercepto representam a heterogeneidade natural entre os indivíduos decorrente de fatores (genéticos ou ambientais) não medidos (MOLENBERGHS e VERBEKE, 2001; DIGGLE *et al.*, 1994).

2.3 Inferência para os modelos

Os coeficientes de regressão do modelo marginal e do modelo de efeitos aleatórios podem ser estimados através do método da máxima verossimilhança. Entretanto, este método produz estimadores viciados para os componentes de variância (DIGGLE, LIANG e ZEGER, 1994). O método da máxima verossimilhança restrita corrige este problema. A expressão da função de verossimilhança restrita é:

$$L_R(\theta) = |X^T H^{-1} X|^{-1/2} \cdot L_V(\theta)$$

em que, θ representa o conjunto de todos os parâmetros no modelo a serem estimados no modelo, X é a matriz $n \times p$ das covariáveis incluídas na análise, $H = \sigma^2 I$ e $L_V(\theta)$ é a função de verossimilhança usual:

$$L_V(\theta) = \prod_{i=1}^n (2\pi)^{-mi/2} |V_i|^{-1/2} (\sigma^2)^{-mi/2} \cdot \exp\left\{-\frac{1}{2\sigma^2} (Y_i - X_i\beta)^T V_i^{-1} (Y_i - X_i\beta)\right\}$$

O estimador de máxima verossimilhança restrita possui as mesmas propriedades que o estimador de máxima verossimilhança. Maiores detalhes sobre este método podem ser encontrados em DIGGLE, LIANG e ZEGER (1994).

No modelo linear de efeitos aleatórios, os coeficientes da regressão são denominados efeitos fixos e o conjunto de variâncias e covariâncias, efeitos aleatórios (GOLDSTEIN, BROWNE e RABASH, 2002). O uso do estimador de máxima verossimilhança restrita (EMVR) produz estimativas não viciadas dos componentes da variância, removendo o vício que existe na estimação da média (DIGGLE *et al.*, 1994). Quando a amostra é grande os EMV e os EMVR são similares mas quando a amostra não é tão grande e o número de regressores (parâmetros) for grande deve ser dada preferência à EMVR.

2.4 Modelo para dados desbalanceados

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 se trabalha com dados balanceados e se pretende avaliar o uso, por exemplo, de um modelo autoregressivo para modelar a covariância dos dados. O modelo autoregressivo

exige que as medidas sejam obtidas em tempos igualmente espaçados, não sendo aplicável para dados desbalanceados (BELLOCO, 2001).

Assim, na análise de dados longitudinais desbalanceados o modelo marginal não pode ser utilizado porque não se consegue estruturar a matriz de variância e covariâncias. A opção recai então no uso do modelo de efeitos aleatórios.

2.5 Interpretação dos parâmetros

A interpretação dos coeficientes de regressão nos modelos de efeitos aleatórios pode ser feito de duas maneiras. Considere por exemplo que a variável resposta seja peso e que uma covariável estado civil, por exemplo (0=solteiro; 1=casado) apresente coeficiente igual a 10,7 unidades (quilos). Uma das interpretações, a “entre indivíduos” indica que a diferença entre dois indivíduos, com uma unidade (casado x solteiro) corresponde a uma mudança de 10,7 quilos no peso (variável resposta). Outra interpretação, a “intra-indivíduos” indica que a mudança de estado civil no mesmo indivíduo (solteiro para casado) corresponderá a uma mudança de 10,7 quilos. Entretanto, a interpretação “real” é uma combinação das duas acima citadas (TWISK, 2003, p. 88).

3 MOTIVAÇÃO CLÍNICA E MÉTODOS

3.1 Hidronefrose fetal

O aparelho urinário é estéril em todo seu trajeto, com exceção do terço terminal da uretra, onde se encontram bactérias oriundas do períneo. Quando microorganismos alcançam outras áreas do trato urinário (via ascendente ou via hematogênica) colonizando e invadindo os tecidos, desencadeia-se a infecção do trato urinário (ITU).

A principal condição de defesa do trato urinário contra invasão, fixação e multiplicação bacteriana é o livre fluxo da urina, desde sua formação no parênquima renal até seu armazenamento temporário na bexiga, seguido da micção. A alteração desse fluxo livre, causada por fatores que a ele se interponham – uropatias congênitas ou outras alterações adquiridas – representa riscos de ITU, quase sempre de caráter recidivante. Por isso, reconhecê-los e tratá-los adequadamente são pontos importantes no manuseio desses pacientes. A investigação do trato urinário por imagens está indicada após o primeiro episódio bem documentado de infecção urinária em qualquer idade e para ambos os gêneros. A indicação dessa investigação está baseada na alta prevalência de malformações congênitas, especialmente o refluxo vesicoureteral, que tem sido identificado em 25% a 50% das crianças avaliadas após um surto de ITU.

Com a utilização crescente da ultra-sonografia obstétrica, aliada à melhoria tecnológica dos equipamentos, as malformações do trato urinário têm sido freqüentemente identificadas intra-útero. Entre essas, a hidronefrose é a alteração mais comumente diagnosticada, sendo relatada uma freqüência de 1:700 gestações. Grande parte delas pode apresentar uma resolução espontânea intra-útero ou no período neonatal, considerando-se, nesse caso, a característica da micção fetal, efetuada em ambiente fechado e de maior pressão – bolsa amniótica. Entretanto as que persistem podem ser indício de graves alterações, obstrutivas ou não, do trato urinário. É de muita importância a avaliação do líquido amniótico pela ultra-sonografia fetal. A diminuição dele pode significar uma baixa taxa de filtração glomerular fetal – como ocorre nas obstruções baixas (válvula de uretra posterior) ou altas (ureterais bilaterais), que indicam pior prognóstico funcional imediato. Dados da Unidade de Nefrologia Pediátrica do Hospital das Clínicas da UFMG (UNP – HC-UFMG) demonstram que as

causas mais freqüentes, em 222 crianças portadoras de hidronefrose fetal persistente são: estenose pieloureteral (26%), hidronefrose idiopática (22%), refluxo vesicoureteral primário (15%), rim multicístico (13%), megaureter primário (10%) e válvula de uretra posterior (8%). A conduta no pós-natal imediato, visa à profilaxia da ITU. De uma maneira geral, entre quatro e sete dias de vida, quando há incremento do ritmo de filtração glomerular do recém-nascido, reduzindo-se a freqüência de resultados falso-negativos, faz-se a ultra-sonografia dinâmica do trato urinário. Conforme o resultado desse exame, outros exames como Uretrocistografia Miccional (UCM) e estudo radiológico da coluna lombossacra são solicitados. Em determinadas situações a propedêutica pode ser complementada com a realização de cintilografia renal estática e dinâmica e, em alguns raros casos, com a urografia excretora. O tratamento adotado vai depender da causa básica encontrada. O diagnóstico precoce tem permitido uma conduta conservadora na maioria dos casos, sendo a correção cirúrgica indicada nos casos de obstrução grave do fluxo urinário. O seguimento deve ser a longo prazo, mantendo-se quimioprofilaxia e avaliação periódica de imagens. Com esse protocolo, tem-se obtido sucesso na prevenção de infecções urinárias, sendo que apenas cinco das 222 crianças da casuística acima referida apresentaram infecções urinárias recidivantes, apesar da quimioprofilaxia. No entanto, nessa mesma casuística, 12% das crianças apresentaram um mau prognóstico, evoluindo para o óbito e/ou falência renal, apesar de todas as intervenções realizadas.

Assim, a presença de hidronefrose fetal deve ser sempre investigada no período neonatal e mantendo-se a urina estéril e um fluxo urinário adequado (SILVA *et al.*, 2005). Definição de hidronefrose pelo MEDLINE é “Aumento anormal do rim ou inchaço devido à dilatação dos cálices renais e pelve renal. Está frequentemente associado com obstrução do ureter ou doença renal crônica que dificultam o fluxo urinário normal da urina até a bexiga”.

A medida do diâmetro antero-posterior da pelve renal (DAP) apresenta sensibilidade adequada para identificar fetos com uropatia significativa bem como aqueles que precisarão de intervenção após o nascimento (BOUZADA *et al.*, 2004). A presença de oligohidrânio e megacistos no ultrassom pré-natal é altamente preditor de obstrução fetal uretral. A obstrução da junção ureteropélvica também pode ser causa de hidronefrose fetal e em 77 pacientes a intervenção cirúrgica foi benéfica naqueles com hidronefrose intensa e função renal diminuída por permitir a melhora ou preservação do parênquima renal. Entretanto, nos pacientes com dilatação moderada ou leve o

tratamento conservador e seguimento clínico foram satisfatórios (APOCALYPSE *et al.*, 2003). Já foi demonstrada correlação entre o comprimento de rim displásico multicístico e parâmetros corporais como peso e altura com “modelo linear de efeito misto” (*linear mixed effect model*) (RABELO *et al.*, 2006). A involução do rim afetado foi mais intensa nos primeiros 30 meses com hipertrofia renal compensatória na unidade renal contralateral mais intensa nos primeiros 24 meses de idade (RABELO *et al.*, 2005-C). Variáveis associadas com involução são gênero, rim não palpável, comprimento renal à admissão (RABELO *et al.*, 2005-B; RABELO *et al.*, 2005-A).

3.2 Coleta de dados

Os dados foram coletados prospectivamente na Unidade de Nefrologia Pediátrica do HC-UFMG pelo grupo de estudo de Hidronefrose fetal e Anomalias Congênitas do Trato Urinário. Todas as crianças tiveram uma avaliação basal intra-útero, outra na primeira semana após o nascimento e em seguida foram avaliadas em média a cada seis meses. O banco de dados original era composto por 241 crianças. Para assegurar a independência entre as medidas foi considerada apenas uma unidade renal, a com maior valor de volume renal.

3.3 Critérios de inclusão

Pacientes com duas ou mais medidas ultra-sonográficas do DAP da pelve.

3.4 Critérios de exclusão

Pacientes submetidos a heminefrectomia

Pacientes com ureterocele (n=5), rim hipoplásico (n=6), válvula de uretra posterior (n=3), rim em ferradura (n=1) e duplicação renal (n=3).

Assim, foram consideradas no banco de dados final 165 unidades renais com um total de 747 observações.

3.5 Variáveis consideradas no presente estudo

A variável resposta, volume renal, é uma variável contínua obtida a partir de uma fórmula baseada em três outras medidas. O banco de dados conta ainda com outras seis variáveis independentes sendo duas contínuas e que se modificam com o tempo, diâmetro antero-posterior (DAP) e superfície corporal (SC) e três dicotômicas que não se modificam com o tempo, prematuro, infecção do trato urinário (ITU) e uropatia. A informação sobre tratamento cirúrgico, é uma variável dicotômica, que muda com o tempo, ou seja os pacientes são classificados como “0” até a data em que são submetidos a intervenção cirúrgica quando passam a ser identificados como “1” (Tabela 1).

A escolha da superfície corporal foi feita pelo fato dela incluir medidas de peso e altura da criança e refletir assim o crescimento da criança. Optou-se por não utilizar índices antropométricos (peso/idade, altura/idade e peso/altura) porque eles refletiriam alterações agudas ou crônicas no estado nutricional (GOULART *et al.* 2005) e para que se contemplassem as duas situações seria necessário o uso de duas variáveis que foram então substituídas por uma única.

TABELA 1
Variáveis do banco de dados

| Código | Significado | Classificação | Muda com tempo | Fórmula |
|----------------------|--|---------------|----------------|-----------------------------------|
| Dependente | | | | |
| Volume | Volume renal (cm ³) | Contínua | Sim | $(RL*RT*RAP*0,523)/1000$ |
| Independentes | | | | |
| DAP | Diâmetro antero-posterior (cm) | Contínua | Sim | |
| SC | Superfície corporal (m ²) | Contínua | Sim | $\sqrt{altura(cm)*peso(kg)/3600}$ |
| Prematuro | Prematuridade (idade gestacional menor que 37 semanas) | Dicotômica | Não | |
| ITU | Ocorrência de infecção do trato urinário | Dicotômica | Não | |
| Uropatia | Presença ou não de uropatia | Dicotômica | Não | |
| Data cirurgia | Identifica medidas que foram realizadas após intervenção cirúrgica | Dicotômica | Sim | |

RL = medida longitudinal do rim

RT = medida transversal do rim

RAP = medida antero-posterior do rim

3.6 Análise estatística

Foi utilizado o software STATA versão 9.5 na análise estatística e para manipulação do banco de dados foi utilizado o SPSS versão 13. A utilização do SPSS foi necessária porque para se calcular a variação entre duas medidas para o mesmo indivíduo (delta) foi necessário que os dados estivessem no formato “largo” (onde cada indivíduo tem um único registro e as medidas são colocadas em diversas colunas). Já para a análise do STATA os dados devem estar no formato “longo” (onde cada indivíduo possui tantos registros quantas forem as medidas ao longo do tempo) (TWISK, 2003).

Os comandos utilizados no SPSS e STATA são apresentados nos Anexos 1 a 3.

Conforme citado na Seção 2 e como os dados do presente estudo são desbalanceados a análise estatística foi realizada utilizando o modelo de efeitos aleatórios considerando-se efeito aleatório no intercepto e no tempo.

Para todas as estratégias (ver Capítulo 4) foi realizada uma análise considerando a variável repostada (volume renal), idade (tempo) e cada covariável sempre com efeito aleatório no intercepto e no tempo, chamada nesta dissertação de análise separada. A seguir, as variáveis que apresentaram valor de p igual ou menor que 0,25 na etapa descrita acima (análise separada) foram incluídas num modelo conjunto. As variáveis foram retiradas então considerando-se o maior valor de p até que restassem apenas variáveis com valor- p menor ou igual a 0,05.

Por último, o modelo com as covariáveis que apresentaram significância estatística e efeito aleatório no intercepto e na idade (tempo) foi comparado ao modelo sem efeito aleatório no tempo para decidir-se pela significância ou não do efeito aleatório no tempo. Como citado anteriormente, ao se retirar o efeito aleatório, “retiram-se” dois componentes estimados do modelo: a variância das inclinações e a covariância entre intercepto aleatório e inclinação aleatória. Por isso, a comparação entre o modelo com e sem efeito aleatório foi feito através do teste da razão de verossimilhança.

4 ESTRATÉGIAS NA ANÁLISE DE DADOS LONGITUDINAIS DESBALANCEADOS

4.1 Introdução

O objetivo desta dissertação é propor diferentes formas de tratamento para dados longitudinais desbalanceados. Inicialmente serão apresentadas estratégias de análise para dados balanceados e posteriormente adaptá-las para desbalanceados. Para uma discussão sobre estratégias de modelagem em dados balanceados sugere-se uma consulta a VIEIRA, BASTOS e HIPPERT (2007).

4.2 Dados balanceados

Na Tabela 2 são apresentadas as estratégias de análise para dados balanceados. Para melhor compreensão destas estratégias são apresentados dados hipotéticos, com apenas uma variável e uma covariável (chamada de “análise separada” nesta dissertação) na Tabela 3.

TABELA 2 – Estratégias de análise para dados balanceados.

| Estratégia | Fórmula | Comentários sobre modelo |
|--|---|--|
| 1 Modelo de efeitos aleatórios | $Peso_{it} = \alpha + \beta altura_{it} + \varepsilon_{ij}$ | Influência da altura no peso ao longo do tempo. Ao colocar t como covariável o tempo pode ser diferente entre as variáveis. |
| 2 Modelo com variáveis de mudança nas covariáveis | $Peso_{it} = \alpha + \beta(\Delta altura)_t + \varepsilon_{ij}$ | Influência da variação da altura (em dois momentos) no peso ao longo do tempo. O peso no tempo t2 será função da mudança na altura de t2 a t1. |
| 3 Somente mudança na variável independente | $\Delta peso_{it} = \beta altura_{it} + \varepsilon_{ii}$ | A variação no peso, e não a medida do peso, será consequência da altura ao longo do tempo |
| 4 Mudança na variável independente e nas covariáveis | $\Delta peso_{it} = \beta(\Delta altura_{it}) + \varepsilon_{ij}$ | A variação no peso, e não a medida do peso, será consequência da variação da altura ao longo do tempo |

TABELA 3 – Dados hipotéticos para exemplificar estratégias de análise

| Indivíduo | Tempo | Resposta (peso) | Independente Altura | Δ altura | Δ peso |
|-----------|-------|--------------------|------------------------|-----------------|---------------|
| 1 | 1 | 15 | 90 | ... | ... |
| 1 | 2 | 18 | 90 | 0 | 3 |
| 1 | 3 | 19 | 91 | 1 | 1 |
| 1 | 4 | 20 | 91 | 0 | 1 |
| 2 | 1 | 13 | 101 | ... | ... |
| 2 | 2 | 15 | 101 | 0 | 2 |
| 2 | 3 | 17 | 101 | 0 | 2 |
| 2 | 4 | 18 | 103 | 2 | 1 |
| 3 | 1 | 9 | 70 | ... | ... |
| 3 | 2 | 9 | 71 | 1 | 0 |
| 3 | 3 | 8 | 72 | 1 | 1 |
| 3 | 4 | 9 | 73 | 1 | 1 |

Podemos observar que os três indivíduos hipotéticos possuem quatro medidas cada e todas obtidas nos mesmos tempos o que caracteriza dados balanceados. Neste exemplo o objetivo é explicar a variação do peso pela variação da altura ao longo do tempo.

A estratégia 1 consiste em ajustar o modelo de efeitos aleatórios padrão em que o peso de um indivíduo ao longo do tempo (coluna 3 da Tabela 3) é explicado pela sua altura também ao longo do tempo (coluna 4 da Tabela 3).

Na estratégia 2 o peso ao longo do tempo (coluna 3 da Tabela 3) é predito pela variação da altura entre o momento atual e momento anterior (coluna 5 da Tabela 3). Nas estratégias 3 e 4 o objetivo é explicar a variação de peso (coluna 6 da Tabela 3). Na estratégia 3 a predição é baseada na altura (coluna 4 da Tabela 3) e na estratégia 4 na variação de altura (coluna 5 da Tabela 3).

4.3 Dados desbalanceados

Na Tabela 4 são apresentadas as mesmas estratégias da Tabela 2 porém com comentários sobre as implicações de aplicá-las na análise de dados desbalanceados. Como citado anteriormente o objetivo é explicar a evolução do volume renal (que caracteriza a hidronefrose) medido repetidamente ao longo do tempo. Entretanto, o número de medidas por indivíduo e o intervalo entre as medidas não é o mesmo o que caracteriza dados desbalanceados.

A estratégia 1, modelo de efeitos aleatórios, não apresenta dificuldades de aplicação em dados desbalanceados. Já na estratégia 2 a variação do volume renal deve ser ajustada de alguma forma pois o intervalo entre as medidas não é uniforme. A

solução utilizada no presente estudo foi padronizar estas variações pelo comprimento do intervalo de tempo, ou seja, considerando-se o indivíduo 1 temos: $(\text{peso2} - \text{peso1}) / (\text{tempo2} - \text{tempo1}) = 3/1 = 3$. Nas estratégias 3 e 4 tanto a variável dependente como as variáveis preditoras devem ser padronizadas pelo comprimento do intervalo de tempo.

TABELA 4 Estratégias de análise para dados desbalanceados

| Fórmula* | Dificuldade em dados desbalanceados |
|---|---|
| 1 $Volumerenal = \beta_0 + \beta_1 \text{tempo} + \beta_2 DAP$ | Sem problemas |
| 2 $Volumerenal = \beta_0 + \beta_1 \text{tempo} + \beta_2 (\Delta DAP)$ | Como o intervalo do tempo é diferente entre as medidas uma solução será a padronização, ou seja, dividir a variação entre dois momentos pelo tempo transcorrido entre eles. |
| 3 $\Delta Volumerenal = \beta_0 + \beta_1 \text{tempo} + \beta_2 DAP$ | Usar padronização pelo intervalo do tempo como descrito na estratégia 2 |
| 4 $\Delta Volumerenal = \beta_0 + \beta_1 (\Delta \text{tempo}) + \beta_2 (\Delta DAP)$ | Usar padronização pelo intervalo do tempo como descrito na estratégia 2 |

* em todos os modelos as variáveis $\beta_{\text{prematuro}} + \beta_{\text{ITU}} + \beta_{\text{uropatia}} + \beta_{\text{momentocirurgia}}$ estarão presentes mas não são apresentadas porque são constantes. Outra variável, “superfície corporal”, será considerada da mesma forma que a variável DAP: valor ou delta.

Neste estudo, como citado anteriormente, a variável tempo é representada pela idade da criança, ao contrário de estudos em que se procura identificar o efeito histórico (VIEIRA, BASTOS e HIPPERT, 2007).

5 RESULTADOS

5.1 Descrição da amostra

No banco de dados foram incluídos 165 pacientes com um total de 747 medidas o que corresponde a uma média de 4,5 medidas por unidade renal. Na Tabela 5 apresenta-se as medidas de tendência central e dispersão das variáveis contínuas e na Tabela 6 as frequências de variáveis categóricas incluídas no banco de dados. Nesta última Tabela encontra-se também o número de pacientes (unidades renais) segundo o número de medidas.

TABELA 5
Medidas de tendência central e dispersão das variáveis quantitativas

| | Amostra | Mínimo | Máximo | Média | Desvio-padrão |
|------------------------------------|---------|--------|--------|-------|---------------|
| Diâmetro antero-posterior | 747 | 3,0 | 79,4 | 13,00 | 10,62 |
| Variação diâmetro antero-posterior | 579 | -12,97 | 10,06 | -0,05 | 1,27 |
| Volume | 747 | 6,99 | 508,15 | 52,84 | 48,07 |
| Variação volume | 579 | -58,93 | 30,63 | 0,96 | 5,57 |
| Idade meses | 747 | 0,03 | 81,67 | 18,20 | 17,12 |
| Superfície corporal | 747 | 0,13 | 1,02 | 0,45 | ,17 |
| Variação superfície corporal | 579 | -0,04 | ,97 | 0,07 | ,17 |

TABELA 6
Frequência de variáveis categóricas

| Variável | Amostra | % |
|-------------------|---------|------|
| Prematuro | | |
| Sim | 14 | 8,5 |
| Não | 151 | 91,5 |
| ITU | | |
| Sim | 31 | 18,8 |
| Não | 134 | 81,2 |
| Uropata | | |
| Sim | 66 | 40,0 |
| Não | 99 | 60,0 |
| Número de medidas | | |
| 2 | 32 | 19,4 |
| 3 | 37 | 22,4 |
| 4 | 24 | 14,5 |
| 5 | 16 | 9,7 |
| 6 | 24 | 14,5 |
| 7 | 13 | 7,9 |
| 8 | 13 | 7,9 |
| 9 | 3 | 1,8 |
| 10 | 3 | 1,8 |

Na Figura 4 observa-se o gráfico *box-plot* do volume renal e o gráfico de correlação do mesmo com o tempo (idade em meses) cujo coeficiente de correlação de Pearson foi 0,325 com $p < 0,001$. Na figura 5 observam-se as mesmas informações para a variável variação do volume renal (delta) cujo coeficiente de correlação de Pearson foi de -0,109 com $p = 0,009$.

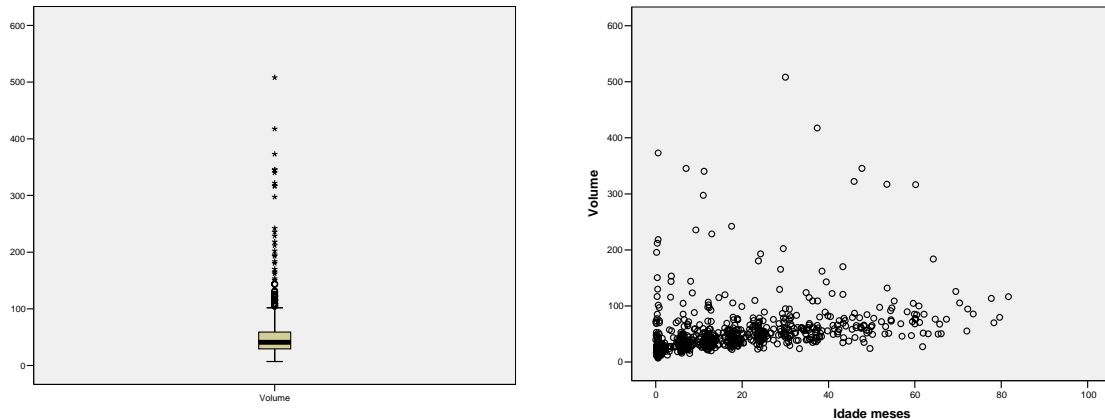


Gráfico *box-plot* do volume renal e gráfico de dispersão deste com a idade (meses)
FIGURA 4

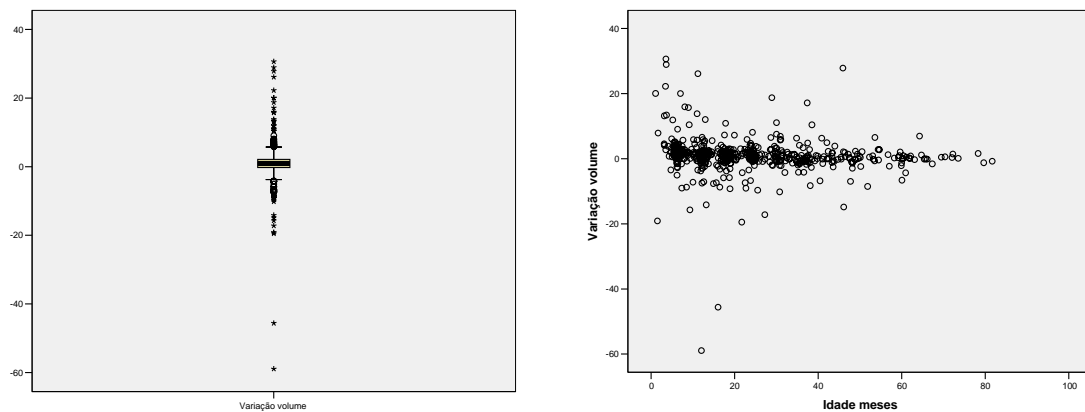


Gráfico *box-plot* da variação do volume renal e gráfico de dispersão deste com a idade (meses)
FIGURA 5

5.2 Estratégia 1 (volume e covariáveis)

Nesta estratégia, inicialmente, foram ajustados 6 modelos considerando-se sempre o volume renal como resposta e efeitos aleatórios no tempo e intercepto além de outra covariável que foi uma das seguintes: diâmetro antero-posterior (DAP), superfície

corporal (SC), prematuro, infecção do trato urinário (ITU), uropatia, data da cirurgia. Na primeira parte da Tabela 7 observa-se o coeficiente, intervalo de confiança 95% e valor-p do tempo, representado nesta dissertação, pela idade em meses. Assim, na primeira parte desta tabela observa-se que o tempo não está associado com o volume renal, quando se considera apenas a covariável superfície corporal (SC). Quando se considera separadamente as outras covariáveis o tempo está sempre associado com o volume renal com um coeficiente em torno de 0,80, ou seja, a variação de um mês está correlacionada com aumento de 0,80 unidades no volume renal.

Na segunda parte da Tabela 7 observa-se os parâmetros (valor-p, coeficiente e intervalo de confiança 95%) das covariáveis. Considerando-se cada uma destas covariáveis separadamente, como já descrito acima, a covariável prematuro não está associada com volume renal e a data da cirurgia apresenta valor-p próximo ao limite da significância estatística. Observa-se ainda, por exemplo, que crianças uropatas tem 36 unidades a mais no volume renal que outras crianças.

Na Tabela 8 encontram-se os modelos de análise conjunta onde todas as variáveis com valor de p igual ou inferior a 0,25 foram incluídas no modelo inicial (modelo 1) e então retiradas uma a uma de acordo com a ausência de significância estatística. Assim, no modelo 1 as variáveis prematuro, ITU e data da cirurgia não apresentaram significância estatística. Apesar do maior valor de p da variável “data da cirurgia”, no modelo 2 optou-se por retirar a variável prematuro porque esta não esteve associada na análise separada. O modelo 3 corresponde à retirada da variável “data da cirurgia” em relação ao modelo 2 e o modelo 4 à retirada da variável ITU em relação ao modelo 3.

TABELA 7
Análise separada para cada covariável e do tempo a com variável resposta

| Covariável | DAP | SC | Prematuro | ITU | Uropatia | Data cirurgia |
|-------------------------------|-------------|---------------|--------------|--------------|---------------|---------------|
| Idade em meses (tempo) | | | | | | |
| Valor-p | <0,001 | 0,59 | <0,001 | <0,001 | <0,001 | <0,001 |
| Coeficiente | 0,89 | 0,12 | 0,78 | 0,78 | 0,78 | 0,81 |
| IC 95% | 0,70 a 1,08 | -0,31 a 0,56 | 0,49 a 1,08 | 0,49 a 1,07 | 0,49 a 1,07 | 0,51 a 1,10 |
| Parâmetros covariável | | | | | | |
| Muda com tempo | Sim | Sim | Não | Não | Não | Sim |
| Codificação | | | 1=sim | 1=sim | 1=sim | 1=sim |
| Valor-p | <0,001 | <0,001 | 0,186 | 0,001 | <0,001 | 0,065 |
| Coeficiente | 3,04 | 59,95 | -13,83 | 23,21 | 35,97 | -12,89 |
| IC 95% | 2,85 a 3,24 | 29,46 a 90,45 | -34,3 a 6,66 | 9,14 a 37,28 | 25,69 a 46,25 | -26,60 a 0,82 |

TABELA 8
Valores de p das covariáveis em modelos de análise conjunta (estratégia 1)

| | Modelo 1 | Modelo 2 | Modelo 3 | Modelo 4 |
|---------------|----------|----------|----------|----------|
| Idade meses | 0,040 | 0,043 | 0,043 | 0,047 |
| DAP | <0,001 | <0,001 | <0,001 | <0,001 |
| SC | <0,001 | <0,001 | <0,001 | <0,001 |
| Prematuro | 0,526 | X | X | X |
| ITU | 0,181 | 0,185 | 0,187 | X |
| Uropatia | 0,002 | 0,002 | 0,001 | 0,003 |
| Data cirurgia | 0,783 | 0,750 | X | X |

X – variável não considerada no modelo por ausência de significância estatística

O modelo final considerando-se efeito aleatório no intercepto e no tempo foi o apresentado na Tabela 9 e apresentou valor do máximo do logaritmo da função de verossimilhança ($\log L(\hat{\theta})$) -3408,35.

TABELA 9
Parâmetros das covariáveis presentes no modelo final

| Covariável | Coefficiente | Erro padrão | Valor-p | IC 95% |
|--------------------------------------|--------------|-------------|---------------|----------------|
| Idade meses (tempo) | 0,31 | 0,15 | 0,047 | 0,004 a 0,61 |
| DAP | 3,19 | 0,11 | <0,001 | 2,98 a 3,41 |
| SC | 57,51 | 11,87 | <0,001 | 34,24 a 80,78 |
| Uropatia | -9,01 | 2,98 | 0,003 | -14,86 a -3,16 |
| Constante | -15,69 | 3,79 | <0,001 | -23,12 a -8,26 |
| Parâmetros de efeitos aleatórios | Estimativa | Erro-padrão | IC 95% | |
| Sd (idade meses) | 0,78 | 0,09 | 0,62 a 0,97 | |
| Sd (constante) | 9,34 | 1,87 | 6,31 a 13,8 | |
| Correlação (idade meses e constante) | 0,33 | 0,28 | -0,27 a 0,75 | |
| Sd (resíduos) | 19,07 | 0,63 | 17,88 a 20,34 | |

Sd = “standard deviation”

Foi testado um novo modelo das mesmas covariáveis apresentadas na Tabela 9 porém sem efeito aleatório no tempo. Este modelo apresentou $\log L(\hat{\theta}) = -3464,09$. O teste de razão de verossimilhança resultou em 112 (valor muito acima do valor crítico) que indica que existe diferença entre os dois modelos e sinaliza que o modelo com efeitos aleatórios no intercepto e no tempo deve ser o escolhido.

O modelo final permite as seguintes interpretações para os parâmetros apresentadas na Tabela 10.

TABELA 10

Interpretação dos parâmetros das variáveis presentes no modelo final da estratégia 1

| Variável | Interpretação |
|-------------|--|
| Idade meses | Aumento de uma unidade na idade é acompanhada de uma mudança de 0,31 unidades no volume renal |
| DAP | Aumento de uma unidade no diâmetro antero-posterior é acompanhada de uma mudança de 3,19 unidades no volume renal |
| SC | Aumento de uma unidade na superfície corporal (m ²) é acompanhada de uma mudança de 57,51 unidades no volume renal |
| Uropatia | Crianças uropatas apresentam um volume renal 9 unidades menor que crianças não uropatas (veja comentário à frente quando são comparadas estratégias 1 e 2) |

Entretanto, o comportamento da variável uropatia, com coeficiente negativo não corresponde ao que seria observado na evolução clínica destes pacientes pois um paciente uropata apresentará maior volume renal pela própria uropatia. Interessante destacar que na análise separada o coeficiente da uropatia foi positivo (35,97) e não negativo) O coeficiente das outras variáveis (DAP e SC) é condizente com a evolução clínica esperada: aumento DAP e da SC explicam aumento no volume renal.

Decidiu-se então por fazer a seleção das covariáveis a serem incluídas no modelo através de um método do tipo *forward*, começando com a variável uropatia e incluindo as demais variáveis com valor $p < 0,25$ na análise separada e decidir pela manutenção ou não delas baseando-se em critérios estatísticos e que correspondam também à evolução clínica esperada dos pacientes. Os modelos são apresentados na Tabela 11. Nesta tabela os valores de p e coeficientes de cada variável são apresentados por coluna e nas linhas pode-se observar quais variáveis foram incluídas em determinado modelo. No modelo 1 observa-se que a variável prematuridade não apresenta significância quando considerada conjuntamente com a variável uropatia. O mesmo ocorre no modelo 2 com a variável ITU. No modelo 3, onde se considera a variável data da cirurgia, observa-se que tanto esta quanto aquela apresentam significância estatística e que o coeficiente das duas apresentam valores condizentes com o esperado na evolução clínica: pacientes uropatas apresentam 39,2 unidades a mais de volume renal e após a cirurgia o volume é reduzido em 18,7 unidades.

No modelo 4 observa-se que a inclusão da variável SC no modelo 3 (que já tinha as variáveis uropatia e data da cirurgia) fez com que o tempo deixasse de apresentar

significância estatística, comportamento este também apresentado na análise separada (Tabela 7). Por isso, decidiu-se não incluir esta variável no modelo seguinte.

No modelo 5, a inclusão da variável DAP “provocou” a ausência de significância da data de cirurgia e também a inversão no coeficiente da uropatia que passou a ser negativo. Estas duas modificações não correspondem à expectativa de evolução clínica e por isso não foram consideradas. Optou-se então por eleger como modelo final o modelo 3 no qual são considerados o tempo, a data de cirurgia e a uropatia.

TABELA 11 – Estratégia 1 - análise conjunta – *forward*

| | Idade meses | Uropatia | Prematuro | ITU | Data da cirurgia | Superfície corporal | DAP |
|-----------------|-------------|----------|-----------|-------|------------------|---------------------|--------|
| Modelo 1 | | | | | | | |
| Valor-p | <0,001 | <0,001 | 0,470 | ... | ... | ... | ... |
| Coeficiente | 0,78 | 35,5 | -6,73 | ... | ... | ... | ... |
| Modelo 2 | | | | | | | |
| Valor-p | <0,001 | <0,001 | ... | 0,128 | ... | ... | ... |
| Coeficiente | 0,78 | 33,4 | ... | 10,3 | ... | ... | ... |
| Modelo 3 | | | | | | | |
| Valor-p | <0,001 | <0,001 | ... | ... | 0,007 | ... | ... |
| Coeficiente | 0,82 | 39,2 | ... | ... | -18,7 | ... | ... |
| Modelo 4 | | | | | | | |
| Valor-p | 0,481 | <0,001 | ... | ... | <0,001 | <0,001 | ... |
| Coeficiente | 0,16 | 38,87 | ... | ... | -24,0 | 62,7 | ... |
| Modelo 5 | | | | | | | |
| Valor-p | <0,001 | 0,007 | ... | ... | 0,936 | ... | <0,001 |
| Coeficiente | 0,90 | -8,2 | ... | ... | 0,36 | ... | 3,2 |

Este modelo final, cujos parâmetros são apresentados na Tabela 12, apresentou valor do máximo do logaritmo da função de verossimilhança ($\log L(\hat{\theta})$) -3662,13.

TABELA 12
Parâmetros das covariáveis presentes no modelo final

| Covariável | Coeficiente | Erro padrão | Valor-p | IC 95% |
|--------------------------------------|-------------|-------------|---------------|----------------|
| Idade meses (tempo) | 0,83 | 0,15 | <0,001 | 0,53 a 1,12 |
| Uropatia | 39,22 | 5,53 | <0,001 | 28,37 a 50,07 |
| Data da cirurgia | -18,69 | 6,89 | 0,007 | -32,20 a -5,19 |
| Constante | 23,87 | 3,51 | <0,001 | 16,99 a 30,74 |
| Parâmetros de efeitos aleatórios | Estimativa | Erro-padrão | IC 95% | |
| Sd (idade meses) | 1,36 | 0,14 | 1,11 a 1,67 | |
| Sd (constante) | 30,0 | 2,38 | 25,67 a 35,05 | |
| Correlação (idade meses e constante) | 0,005 | 0,16 | -0,30 a 0,30 | |
| Sd (resíduos) | 22,87 | 0,79 | 21,38 a 24,47 | |

Sd = “standard deviation”

Outro modelo, com as mesmas covariáveis apresentadas na Tabela 12, porém sem efeito aleatório no tempo apresentou $\log L(\hat{\theta}) = -3838,43$. O teste de razão de verossimilhança resultou em 352 (valor muito acima do valor crítico) que indica que existe diferença entre os dois modelos e sinaliza que o modelo com efeitos aleatórios no intercepto e no tempo deve ser o escolhido.

Assim, o modelo final da estratégia 1 indica que o volume renal médio no início do acompanhamento, para não uropatas que não haviam feito cirurgia (categorias de referência consideradas para as covariáveis) é de 23,87 cm³, aumenta 0,83 cm³ a cada mês e que os uropatas apresentam em média 39 cm³ a mais que os não uropatas e que após a cirurgia o volume renal dos uropatas reduz em média 18,7cm³.

5.3 Estratégia 2 (variável dependente sem delta e covariáveis com delta)

Da mesma forma que na estratégia 1, nesta segunda estratégia, inicialmente, foram ajustados 6 modelos considerando-se sempre o volume renal como resposta e efeitos aleatórios no tempo e intercepto além de outra covariável que foi uma das seguintes: variação no diâmetro antero-posterior (delta DAP), variação da superfície corporal (delta SC), prematuro, infecção do trato urinário (ITU), uropatia, data da cirurgia. Importante destacar que a Tabela 13 é diferente da Tabela 7 somente com relação aos parâmetros das variáveis com delta (delta DAP e delta SC) pois as demais covariáveis e variável resposta são exatamente as mesmas nas duas estratégias e por consequência nas duas tabelas.

São apresentados na primeira parte da Tabela 13 o coeficiente, intervalo de confiança 95% e valor-p do tempo, representado nesta dissertação, pela idade em meses. Assim, observa-se que o tempo está associado com o volume renal quando se considera cada uma das variáveis separadamente. Quando são consideradas as duas primeiras variáveis, as que apresentam delta, os valores estimados para o coeficiente do tempo ficaram em torno de 0,50, o que indica que a mudança de uma unidade no tempo (um mês) corresponde a um aumento de 0,50 no volume renal. Considerando as demais covariáveis, o coeficiente estimado para o tempo ficou em torno de 0,80.

Na segunda parte da Tabela 13 observa-se os parâmetros (valor-p, coeficiente e intervalo de confiança 95%) das covariáveis. Considerando-se cada uma destas

covariáveis separadamente, como já descrito acima, prematuro não está associada com volume renal e a data da cirurgia apresenta valor-p próximo ao limite da significância estatística (assim como na estratégia 1). A principal diferença entre as Tabelas 7 e 13 é que na Tabela 7, quando se considera a SC, o tempo não apresenta associação com volume renal (mas SC apresenta) e na Tabela 13 quando se considera a covariável delta SC o tempo está associado (mas a covariável delta SC não está).

TABELA 13

Análise separada para cada covariável com delta e do tempo com variável resposta

| Covariável | Delta DAP | Delta SC | Prematuro | ITU | Uropatia | Data cirurgia |
|-------------------------------|-------------|---------------|--------------|--------------|---------------|---------------|
| Idade em meses (tempo) | | | | | | |
| Valor-p | <0,001 | <0,001 | <0,001 | <0,001 | <0,001 | <0,001 |
| Coefficiente | 0,51 | 0,53 | 0,78 | 0,78 | 0,78 | 0,81 |
| IC 95% | 0,27 a 0,76 | 0,24 a 0,82 | 0,49 a 1,08 | 0,49 a 1,07 | 0,49 a 1,07 | 0,51 a 1,10 |
| Parâmetros covariável | | | | | | |
| Muda com tempo | Sim | Sim | Não | Não | Não | Sim |
| Codificação | | | 1=sim | 1=sim | 1=sim | 1=sim |
| Valor-p | <0,001 | 0,60 | 0,186 | 0,001 | <0,001 | 0,065 |
| Coefficiente | 7,81 | 3,44 | -13,83 | 23,21 | 35,97 | -12,89 |
| IC 95% | 6,33 a 9,29 | -9,38 a 16,26 | -34,3 a 6,66 | 9,14 a 37,28 | 25,69 a 46,25 | -26,60 a 0,82 |

Na Tabela 14 encontram-se os modelos de análise conjunta onde todas as variáveis com valor de p igual ou inferior a 0,25 foram incluídas no modelo inicial (modelo 1) e então retiradas uma a uma de acordo com a ausência de significância estatística. Assim, no modelo 1, no qual a variável delta SC não foi nem considerada, as variáveis prematuro e ITU não apresentaram significância estatística. Estas variáveis foram retiradas respectivamente nos modelos 2 e 3 restando então as variáveis tempo (idade meses), variação do DAP, uropatia e data em que ocorreu a cirurgia cujos parâmetros são apresentados na Tabela 11. O modelo final apresentou $\log L(\hat{\theta}) = -2791,64$.

TABELA 14

Valores de p das covariáveis em modelos de análise conjunta (estratégia 2)

| | Modelo 1 | Modelo 2 | Modelo 3 |
|---------------|----------|----------|----------|
| Idade meses | <0,001 | <0,001 | <0,001 |
| Delta DAP | <0,001 | <0,001 | <0,001 |
| Delta SC | X | X | X |
| Prematuro | 0,556 | X | X |
| ITU | 0,176 | 0,182 | X |
| Uropatia | <0,001 | <0,001 | <0,001 |
| Data cirurgia | 0,036 | 0,032 | 0,024 |

Interessante observar que mesmo a variável delta DAP tendo sido padronizada pelo tempo a variável idade em meses foi significativa. O modelo com efeito aleatório apenas no intercepto (mas não no tempo) apresentou $\log L(\hat{\theta}) = -2825,87$, bem inferior ao valor do modelo com efeito aleatório também no tempo indicando que os dois modelos são diferentes confirmando assim a importância da variável idade meses e que o modelo com efeito aleatório no tempo deve ser considerado.

TABELA 15
Parâmetros das covariáveis presentes no modelo final

| Covariável | Coefficiente | Erro padrão | Valor-p | IC 95% |
|--------------------------------------|--------------|-------------|---------|---------------|
| Idade meses (tempo) | 0,58 | 0,12 | <0,001 | 0,33 a 0,82 |
| Delta DAP | 7,49 | 0,75 | <0,001 | 6,02 a 8,96 |
| Uropatia | 39,29 | 6,09 | <0,001 | 27,36 a 51,22 |
| Data da cirurgia | -16,75 | 7,43 | 0,024 | -31,3 a -2,19 |
| Constante | 29,85 | 3,82 | <0,001 | 22,37 a 37,33 |
| Parâmetros de efeitos aleatórios | Estimativa | Erro-padrão | | IC 95% |
| Sd (idade meses) | 0,98 | 0,13 | | 0,76 a 1,26 |
| Sd (constante) | 30,04 | 2,91 | | 24,84 a 36,34 |
| Correlação (idade meses e constante) | 0,44 | 0,16 | | 0,08 a 0,71 |
| Sd (resíduos) | 19,76 | 0,81 | | 18,22 a 21,42 |

Assim, o modelo final da segunda estratégia indica que a média do volume renal no início do acompanhamento é de $29,85\text{cm}^3$ (considerando as categorias de referência das demais covariáveis), com aumento de $0,58\text{cm}^3$ a cada mês. Os uropatas apresentam $39,3\text{cm}^3$ a mais que os não-uropatas e que se reduz em $16,8\text{cm}^3$ após a cirurgia. A mudança de uma unidade na variação do DAP (deltaDAP) explica um aumento de $7,49\text{cm}^3$ no volume renal.

Para entender melhor a associação de deltaDAP com volume renal podemos exemplificar que caso uma criança apresente aumento de um cm no DAP nos primeiros três meses de vida e entre o terceiro e quarto mês de vida apresente aumento de dois cm (aumento na variação do comprimento do DAP de um) terá um aumento no volume renal esperado pelo modelo de $7,5\text{cm}^3$ entre o terceiro e quarto mês. Ou seja, o delta DAP pode ser interpretado como aumento na velocidade de mudança entre os DAPs.

5.4 Comparação da estratégia 1 com estratégia 2

As estratégias 1 e 2 já apresentadas identificam covariáveis com a finalidade de explicar a mesma variável resposta: volume renal. A diferença entre estas estratégias é que na segunda as informações sobre DAP e SC são trabalhadas como variações (deltas), ou seja, como a variação delas entre dois momentos no mesmo indivíduo.

A Tabela 16 apresenta um resumo dos parâmetros das covariáveis observadas nos modelos finais de análise conjunta das duas estratégias (já apresentados nas Tabelas 12 e 15, respectivamente).

Pode-se observar que os resultados são muito semelhantes entre as duas estratégias com as mesmas variáveis (exceção de deltaDAP que foi avaliada apenas na segunda estratégia) e com coeficientes muito próximos. Importante frisar porém que caso a modelagem tivesse sido baseada apenas nos critérios de significância estatística o modelo final teria sido o apresentado na tabela 9 onde os uropatas apresentariam em média volume renal menor que os não-uropatas.

TABELA 16
Resumo dos parâmetros das covariáveis presentes no modelo final nas estratégias 1 e 2

| Covariáveis | Estratégia 1 | | Estratégia 2 | |
|---------------------|--------------|----------------|--------------|---------------|
| | Coefficiente | IC 95% | Coefficiente | IC 95% |
| Idade meses (tempo) | 0,83 | 0,53 a 1,12 | 0,58 | 0,33 a 0,82 |
| Uropatia | 39,22 | 28,37 a 50,07 | 39,29 | 27,36 a 51,22 |
| Data da cirurgia | -18,69 | -32,20 a -5,19 | -16,75 | -31,3 a -2,19 |
| Delta DAP | X | X | 7,49 | 6,02 a 8,96 |
| Constante | 23,87 | 16,99 a 30,74 | 29,85 | 22,37 a 37,33 |

X – indica ausência de significância estatística

5.5 Estratégia 3 (delta na variável resposta)

A terceira estratégia proposta na presente pesquisa foi explicar a variação da variável resposta (delta) a partir das covariáveis. Entretanto, durante a modelagem com efeito aleatório no intercepto e tempo o STATA apresentou mensagens de erro que inviabilizaram o cálculo dos erros-padrão dos parâmetros estimados para a variável

DAP. Assim, nesta variável foi ajustado o modelo com efeito aleatório somente no intercepto.

Da mesma forma que nas estratégias anteriores nesta estratégia, inicialmente, foram ajustados 6 modelos considerando-se sempre o delta volume renal como resposta e efeitos aleatórios no tempo e intercepto além de outra covariável que foi uma das seguintes: diâmetro antero-posterior (DAP), superfície corporal (SC), prematuro, infecção do trato urinário (ITU), uropatia, data da cirurgia. Na primeira parte da Tabela 17 observa-se o coeficiente, intervalo de confiança 95% e valor-p do tempo, representado nesta dissertação, pela idade em meses. Assim, na primeira parte desta tabela observa-se que o tempo, avaliado pela idade em meses, só não está associado com a variável resposta (delta volume renal) quando se considera a variável superfície corporal. Em todas as outras, o coeficiente permanece negativo com pequeno valor (entre -0,02 e -0,04) indicando que a cada mês ocorre uma variação no volume renal entre -0,02 e -0,04.

Na segunda parte da tabela observa-se os parâmetros (valor-p, coeficiente e IC95%) das diversas covariáveis quando comparadas com a variação do volume renal: prematuro, superfície corporal e uropatia não estão associadas com delta volume renal. A única variável associada separadamente com delta volume renal foi a DAP e cabe lembrar que este modelo foi ajustado sem efeito aleatório no tempo porque quando se considerava este efeito a estimativa dos parâmetros não foi obtida pelos métodos numéricos. As variáveis ITU e data da cirurgia apresentaram valor-p inferior a 0,15 e foram incluídas no modelo de análise conjunta.

TABELA 17

Análise separada para cada covariável e do tempo com delta da variável resposta

| Covariável | DAP* | SC | Prematuro | ITU | Uropatia | Data cirurgia |
|-------------------------------|---------------|---------------|---------------|---------------|---------------|----------------|
| Idade em meses (tempo) | | | | | | |
| Valor-p | 0,004 | 0,450 | 0,010 | 0,012 | 0,012 | 0,028 |
| Coeficiente | -0,04 | -0,02 | -0,04 | -0,04 | -0,04 | -0,03 |
| IC 95% | -0,07 a -0,01 | -0,09 a -0,04 | -0,06 a -0,01 | -0,06 a -0,01 | -0,06 a -0,01 | -0,06 a -0,001 |
| Parâmetros covariável | | | | | | |
| Muda com tempo | Sim | Sim | Não | Não | Não | Sim |
| Codificação | | | 1=sim | 1=sim | 1=sim | 1=sim |
| Valor-p | 0,001 | 0,679 | 0,895 | 0,128 | 0,673 | 0,081 |
| Coeficiente | 0,07 | -1,78 | 0,12 | -0,83 | -0,20 | -1,16 |
| IC 95% | 0,03 a 0,12 | -10,26 a 6,69 | -1,59 a 1,82 | -1,88 a 0,23 | -1,11 a 0,72 | -2,46 a 0,14 |

*modelo com efeito aleatório somente no intercepto

Na Tabela 18 encontra-se o modelo onde todas as variáveis com valor de p igual ou inferior a 0,25 foram incluídas (modelo 1). Nos demais modelos as covariáveis

foram retiradas uma a uma de acordo com a ausência de significância estatística. Assim, no modelo 1, não foram consideradas as variáveis SC, prematuro e uropatia. A variável ITU apresentou valor-p sem significância estatística e foi retirada no segundo modelo considerado modelo final que apresentou $\log L(\hat{\theta}) -1806,34$.

TABELA 18
Valores de p das covariáveis em modelos de análise conjunta (estratégia 3)

| | Modelo 1 | Modelo 2 |
|---------------|----------|----------|
| Idade meses | 0,032 | <0,029 |
| DAP | <0,001 | <0,001 |
| SC | X | X |
| Prematuro | X | X |
| ITU | 0,088 | X |
| Uropatia | X | X |
| Data cirurgia | 0,001 | <0,001 |

Os coeficientes estimados são apresentados na Tabela 19, onde pode ser observado que a variação do volume renal fica menor em 0,03 unidades a cada mês, aumenta 0,11 unidades para cada variação no DAP e reduz 2,54 unidades após a cirurgia (considerando as categorias de referência das demais covariáveis).

TABELA 19
Parâmetros das covariáveis presentes no modelo final (estratégia 3)

| Covariável | Coefficiente | Erro padrão | Valor-p | IC 95% |
|---------------------|--------------|-------------|---------|----------------|
| Idade meses (tempo) | -0,03 | 0,01 | 0,03 | -0,06 a -0,001 |
| DAP | 0,11 | 0,02 | <0,001 | 0,06 a 0,16 |
| Data da cirurgia | -2,54 | 0,72 | <0,001 | -3,95 a -1,13 |
| Constante | 0,64 | 0,47 | 0,18 | -0,29 a 1,57 |

5.6 Estratégia 4 (delta na variável dependente e nas covariáveis)

A quarta estratégia proposta na presente pesquisa foi explicar a variação da variável resposta (delta) pelas covariáveis utilizando as covariáveis delta DAP e delta SC. Entretanto, no ajuste com o modelo de efeito aleatório no intercepto e tempo, o STATA apresentou mensagens de erro que inviabilizaram o cálculo dos erros-padrão

dos parâmetros estimados para a variável DeltaDAP. Assim, foi ajustado o modelo com efeito aleatório somente no intercepto como aconteceu na estratégia 3.

A tabela 20 contém a análise separada dos modelos de cada uma das covariáveis com a variável resposta e tempo. Observa-se na primeira parte da Tabela 20 que o tempo, avaliado pela idade em meses, está associado com a variável resposta (delta volume renal) em todos os modelos, quando se considera cada uma das covariáveis com coeficiente de -0,04 o que indica que a variação do volume renal é reduzida em aproximadamente 0,04 unidades com o tempo (mês).

Na segunda parte da Tabela 20 observa-se os parâmetros (valor-p, coeficiente e intervalo de confiança) das covariáveis: prematuro, delta da superfície corporal e uropatia não estão associadas com volume renal. As variáveis ITU e data da cirurgia apresentaram valor-p inferior a 0,15 e foram incluídas no modelo inicial de análise conjunta. A única variável associada com alteração no volume renal foi o delta DAP.

TABELA 20

Análise separada para cada covariável e do tempo com delta da variável resposta

| Covariável | DeltaDAP* | DeltaSC | Prematuro | ITU | Uropatia | Data cirurgia |
|-------------------------------|---------------|---------------|---------------|---------------|---------------|----------------|
| Idade em meses (tempo) | | | | | | |
| Valor-p | <0,001 | 0,008 | 0,010 | 0,012 | 0,012 | 0,028 |
| Coeficiente | -0,04 | -0,04 | -0,04 | -0,04 | -0,04 | -0,03 |
| IC 95% | -0,06 a -0,02 | -0,07 a -0,01 | -0,06 a -0,01 | -0,06 a -0,01 | -0,06 a -0,01 | -0,06 a -0,001 |
| Parâmetros covariável | | | | | | |
| Muda com tempo | Sim | Sim | Não | Não | Não | Sim |
| Codificação | | | 1=sim | 1=sim | 1=sim | 1=sim |
| Valor-p | <0,001 | 0,564 | 0,895 | 0,128 | 0,673 | 0,081 |
| Coeficiente | 2,70 | 0,829 | 0,12 | -0,83 | -0,20 | -1,16 |
| IC 95% | 2,42 a 2,98 | -1,99 a 3,65 | -1,59 a 1,82 | -1,88 a 0,23 | -1,11 a 0,72 | -2,46 a 0,14 |

*modelo com efeito aleatório somente no intercepto

Na Tabela 21 encontram-se os passos para obtenção do modelo final onde todas as variáveis com valor de p igual ou inferior a 0,25 foram incluídas no modelo inicial (modelo 1). Neste modelo as variáveis ITU e data cirurgia não apresentaram significância estatística e foram testados dois modelos com a retirada de cada uma delas (modelos 2A e 2B) e nenhuma delas apresentou significância estatística. Assim, o modelo final é o modelo 3 que inclui somente delta DAP, cujos parâmetros são apresentados na Tabela 22 e que apresentou $\log L(\hat{\theta}) -1675,78$.

TABELA 21
Valores de p das covariáveis em modelos de análise conjunta (estratégia 4)

| | Modelo 1 | Modelo 2A | Modelo 2B | Modelo 3 |
|---------------|----------|-----------|-----------|----------|
| Idade meses | <0,001 | <0,001 | <0,001 | <0,001 |
| DeltaDAP | <0,001 | <0,001 | <0,001 | <0,001 |
| DeltaSC | X | X | X | X |
| Prematuro | X | X | X | X |
| ITU | 0,616 | X | 0,632 | X |
| Uropatia | X | X | X | X |
| Data cirurgia | 0,851 | 0,913 | X | X |

TABELA 22
Parâmetros das covariáveis presentes no modelo final (estratégia 4)

| Covariável | Coefficiente | Erro padrão | Valor-p | IC 95% |
|---------------------|--------------|-------------|---------|---------------|
| Idade meses (tempo) | -0,04 | 0,01 | 0,001 | -0,06 a -0,02 |
| Delta DAP | 2,70 | 0,14 | <0,001 | 2,42 a 2,98 |
| Constante | 1,95 | 0,31 | <0,001 | 1,35 a 2,56 |

Assim, pode-se interpretar que a variação do volume renal fica menor em 0,04 unidades a cada mês, e que a mudança na velocidade de variação o DAP (delta DAP) de uma unidade explica o aumento de 2,7 unidades na velocidade de variação do volume renal (delta volume).

5.7 Comparação da estratégia 3 com estratégia 4

As estratégias 3 e 4 já apresentadas identificam covariáveis com a finalidade de explicar a mesma variável resposta: variação do volume renal, identificada como delta volume. A diferença básica entre estas estratégias é que na quarta estratégia as informações sobre DAP e SC são trabalhadas como variações (deltas), ou seja, como a variação delas entre dois momentos está relacionada com a variação do volume renal entre estes momentos.

A Tabela 23 apresenta um resumo dos parâmetros das covariáveis observadas nos modelos finais de análise conjunta das duas estratégias já apresentados e que permite observar que o tempo apresenta coeficientes muito próximos e que a constante

apresenta o mesmo sinal positivo nestas estratégias. A principal mudança é a exclusão da variável data da cirurgia quando se considera a variação do DAP.

TABELA 23

Resumo dos parâmetros das covariáveis presentes no modelo final nas estratégias 1 e 2

| Covariáveis | Estratégia 3 | | Estratégia 4 | |
|---------------------|--------------|----------------|--------------|---------------|
| | Coeficiente | IC 95% | Coeficiente | IC 95% |
| Constante | 0,64 | -0,29 a 1,57 | 1,96 | 1,35 a 2,56 |
| Idade meses (tempo) | -0,03 | -0,06 a -0,001 | -0,04 | -0,06 a -0,02 |
| DAP | 0,11 | 0,06 a 0,16 | X | X |
| Data da cirurgia | -2,54 | -3,95 a -1,13 | X | X |
| Delta DAP | X | X | 2,70 | 2,42 a 2,98 |

X – indica ausência de significância estatística

6 CONCLUSÕES

Os ganhos pessoais com a realização deste trabalho foram diversos, desde novas amizades, até novo conhecimento teórico que poderá ser usado em pesquisas na área de saúde. O maior ganho deve ter sido a constatação de algo que não é nenhuma novidade: a análise estatística dos dados deve ser feita em conjunto com a interpretação “clínica” no caso da área da saúde. Isto foi particularmente importante na estratégia 1 onde o modelo ajustado somente com critérios estatísticos (significância estatística) não corresponde à expectativa clínica.

Espera-se que uma das principais contribuições desta dissertação seja mostrar que a utilização de estudos longitudinais na área da saúde, onde os dados são frequentemente desbalanceados, é muito importante, exequível e que permite a utilização de diversas estratégias de modelagem aqui apresentadas.

Entretanto, a interpretação do delta na variável resposta (delta volume renal) não é muito simples e deve ser bem discutida com os pesquisadores.

7 REFERÊNCIAS BIBLIOGRÁFICAS

APOCALYPSE GT, OLIVEIRA EA, RABELO EAS, DINIZ JS, MARINO VS, PEREIRA AK, SIMAL CJ, GAZOLLA LP, FAGUNDES TA. Outcome of apparent ureteropelvic obstruction identified by investigation of fetal hydronephrosis. *Int Urol Nephrol* 2003; 35 (4): 441-448.

BARACHO SMLN. Tratamento de dados ausentes em estudos longitudinais. Dissertação de mestrado, Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, 2003.

BELLOCO R. *Analysis of longitudinal data in Stata, Splus and SAS*. Department of Medical Epidemiology, Karolinska Institutet. Stockholm, Sweden. March 12, 2001. E-mail: rino@mep.ki.se

BOUZADA MC, OLIVEIRA EA, PEREIRA AK, LEITE HV, RODRIGUES AM, FAGUNDES LA, GONÇALVES RP, PARREIRAS R. Diagnostic accuracy of fetal renal pelvis anteroposterior diameter as a predictor of uropathy: a prospective study. *Pediatr Radiol*. 2004; 34 (10): 798-804.

DIGGLE PJ, LIANG KY, ZEGER SL. *Analysis of longitudinal data*. Clarendon Press: Oxford, 1994.

EKUMA O, LIX L. *Random effects models for longitudinal data – continuous data*. Manitoba Centre for Health Policy. May, 2004.

FAUSTO MA. Avaliação longitudinal do crescimento de lactentes nascidos de mães vivendo com HIV/AIDS da coorte de Belo Horizonte [tese de Doutorado]. Belo Horizonte: UFMG; 2005.

GOLDSTEIN H, BROWNE W, RABASH J. Multilevel modelling of medical data. *Statist. Med.*, 2002.

GOULART EMA, CORRÊA EJ, LEÃO E, XAVIER CC, ABRANTES MM. Avaliação do crescimento. In: Leão E, Corrêa EJ, Viana MB, Mota JAC. *Pediatria Ambulatorial*. 4ª edição. Belo Horizonte: COOPMED. 2005.

LAIRD NM, WARE JH. Random-effects for Longitudinal Data. *Biometrics*. 1982; 38 (4): 963-974.

LIMA ACL. Modelagem conjunta de dados longitudinais e de sobrevivência [Dissertação]. Belo Horizonte: UFMG; 2007.

MOLENBERGHS G, VERBEKE G. *A Review on Linear Mixed Models for Longitudinal Data, Possibly Subject to Dropout*. 2001. URL: [http:// biostatistik.uibk.ac.at/roes/papers/RoES_2001_Molenberghs_Verbeke_Paper.pdf](http://biostatistik.uibk.ac.at/roes/papers/RoES_2001_Molenberghs_Verbeke_Paper.pdf) (consultado em 30 de julho de 2003)

OLIVEIRA EA, RABELO EA, PEREIRA AK, DINIZ JSS, CABRAL AC, LEITE HV, SILVA JMP, FAGUNDES TA. Prognostic factors in prenatally-detected posterior urethral valves: a multivariate analysis. *Pediatr Surg Int*. 2002; 18 (8): 662-667.

RABELO EAS, OLIVEIRA EA, DINIZ JS, SILVA JMP, FILGUEIRAS MT, PEZZUTI IL, TATSUO ES. Natural history of multicystic kidney conservatively managed: a prospective study. *Pediatr Nephrol.* 2005; 20 (5): 690-691.

RABELO EAS, OLIVEIRA EA, SILVA GS, PEZZUTI IL, TATSUO ES. Predictive factors of ultrasonographic involution of prenatally detected multystic dysplastic kidney. *BJU Int.* 2005; 95 (6): 868-871.

RABELO EAS, OLIVEIRA EA, SILVA JM, BOUZADA MC, SOUSA BC, ALMEIDA MN, TATSUO ES. Conservative management of multicystic dysplastic kidney: clinical course and ultrasound outcome. *Jornal de Pediatria (Rio J)* 2005; 81(5): 400-404.

RABELO EAS, OLIVEIRA EA, SILVA JMP, OLIVEIRA DS, COLOSIMO EA. Ultrasound progression of prenatally detected multicystic dysplastic kidney. *Urology.* 2006; 68 (5): 1098-1102.

SILVA JMP, DINIZ JS, OLIVEIRA EA, CARDOSO LS, MARINO VS, PIMENTA MR, MATOS CC, VIEIRA SB. Features of primary vesicoureteral reflux and renal damage in children at a single institution in Brazil from 1969 to 1999. *Int Urol Nephrol.* 2003; 35 (2): 161-168.

SILVA JMP, DINIZ JSS, OLIVEIRA EA, CARDOSO LSB, MARINO VS. Infecção do trato urinário. In: LEÃO E, CORRÊA EJ, MOTA JAC, VIANA MB. *Pediatria Ambulatorial.* 4ª edição COOPMED. Belo Horizonte. 2005: 600-609

SILVA JMP, DINIZ JSS, MARINO VS, LIMA EM, CARDOSO LS, VASCONCELOS MA, OLIVEIRA EA. Clinical course of 735 children and adolescents with primary vesicoureteralreflux. *Pediatric Nephrology* 2006; 21(7): 981-988.

SILVA JMP, DINIZ JSS, SILVA AC, AZEVEDO MV, PIMENTA MR, OLIVEIRA EA. Predictive factors of chronic kidney disease in severe vesicoureteral reflux. *Pediatric Nephrology.* 2006; 21 (9): 1285-1292.

SILVA JMP, OLIVEIRA EA, DINIZ JS, BOUZADA MC, VERGARA RM, SOUZA BC. Clinical course of prenatally detected primary vesicoureteral reflux. *Pediatric Nephrology* 2006; 21 (1) 86-91.

TWISK JWR. *Applied longitudinal data analysis for epidemiology – a practical guide.* Cambridge University Press. Cambridge. United Kingdom. 2003 (301 p).

VERBEKE G, MOLENBERGHS G. *Linear mixed models for longitudinal data.* Springer. 2000. New York.

VIEIRA MDT, BASTOR RR, HIPPERT HS. Evaluating different longitudinal modelling strategies: an application to BHPS data. In: ISI 2007 – 56th session of the International Statistical Institute, 2007, Lisboa. *Bulletin of the International Statistical Institute 56th Session – Proceedings ISI 2007.* Lisboa: Tziranda, 2007.

ANEXOS

ANEXO 1 -Transferir dados para STATA

1. Abrir planilha no Excell
2. Mudar vírgula para ponto: Iniciar, configurações, painel de controle, opções regionais e de idioma para inglês.
3. Colocar variáveis de grupos como categóricas (texto) para permitir gráficos separados
4. Salvar arquivo do Excell como “texto separado por tabulações”

ANEXO 2 – Comandos do STATA

Para abrir arquivo no STATA

1. File
2. Import
3. ASCII created by spreadsheet
4. Com o “botão” browse selecione o arquivo desejado e mude a opção de seleção para txt
5. Selecione a opção “Tab-delimited data”
6. Digite iis seguido da variável identificadora
7. Digite tis seguido da variável que identifica tempo

Nos comandos abaixo substituir

var_resposta pela variável dependente

var_tempo pela variável que informa o tempo

var_identificadora pela variável que identifica as medidas de um mesmo indivíduo

covariável(is) pela(s) variável(is) desejadas

Comando de efeitos mistos

Comando de efeitos aleatórios no intercepto e no tempo

8. `xtmixed var_resposta var_tempo covariáveis || var_identificadora: var_tempo, ml cov(un)`
9. `xtreg var_resposta var_tempo covariável(is)`

Para fazer gráficos

`sort var_identificadora var_tempo`

`twoway (connected var_resposta var_tempo, connect(ascending)), by(covariável)`
(atenção para vírgulas e parenteses)

ANEXO 3 – Transformado bancos de dados “largos” em “longos” (uso do SPSS)

TRANSFORMAR DADOS LONGITUDINAIS EM COLUNAS (para calcular delta)

- 1) SPSS abre Excell sem problemas
- 2) Menu Data/ Reestructure
- 3) Opção “Reestructure selected cases into variables”

Atenção pois as variáveis que mudam com o tempo serão consideradas como mudança (será criada diversas variáveis)

Para calcular os deltas lembrar de padronizar pelo intervalo do tempo

Deve-se criar uma coluna em branco para o primeiro delta de cada variável pois o delta entre a primeira e segunda observações será usado para explicar a segunda observação e não a primeira. Esta coluna deve ser identificada como delta1 para que fique em branco no banco de dados e não seja considerada na análise.

TRANSFORMAR DADOS DE COLUNAS EM LONGITUDINAIS

- 1) Menu Data, opção “Reestructure Data”
- 2) Usar a opção “Reestructure selected variables into cases”
- 3) Step 2: Escolher a opção “More than one” e informar quantas variáveis
Certifique-se que existem o mesmo número de variáveis para cada grupo (nas de delta colocar pelo menos uma em branco para que todas tenham 10 por exemplo)
- 4) Step3:
 - a. escolher variável identificadora,
 - b. Associar variáveis similares com a transx
 - c. Identificar variáveis fixas

Anote num pedaço de papel quais variáveis foram relacionadas com qual trans para que você possa renomeá-las mais facilmente depois pois elas serão identificadas apenas como transX
- 5) Step4:
 - d. Identifica quantas variáveis identificadoras serão necessárias
- 6) Step5
 - e. Escolher como medidas repetidas serão identificadas
- 7) No step 6 “variables to cases: options” escolha as seguintes opções
 - “keep and treat as fixed variables”
 - “discard the data” (*para que o novo banco de dados considere apenas as linhas completas*)
- 8) Salvar como Excell
- 9) Conferir se variáveis numéricas estão como número
- 10) Salvar como texto separado por tabulações