

## **Efeito da imputação de dados faltantes em banco de dados de infecção de sítio cirúrgico em pacientes ortopédicos em Belo Horizonte.**

Flávia Falci Ercole <sup>1</sup>, Mariângela Carneiro <sup>2</sup>, Tânia Couto M Chianca<sup>1</sup>, Denise Duarte <sup>3</sup>

### **Resumo**

A ocorrência de informações faltantes em banco de dados, principalmente na área da saúde é uma situação comum que limita a análise estatística de dados completos. Foram aplicadas técnicas de imputação para variáveis com valores faltantes em dados de infecção de sítio cirúrgico em cirurgias ortopédicas realizadas em hospitais de Belo Horizonte, Minas Gerais, Brasil. As análises estatísticas feitas usando o software R de domínio público compararam os resultados obtidos com o banco de dados imputado e com o banco de dados completo analisados por Ercole (2006). De acordo com o tipo da variável foi utilizado um ou mais procedimentos de imputação descritos por Little e Rubin (1987) para dados faltantes. Para os dados relativos à infecção de sítio cirúrgico em ortopedia, a imputação foi eficiente e foram produzidos resultados semelhantes aos obtidos com o banco completo.

---

<sup>1</sup> Professor Doutor. Escola de Enfermagem da Universidade Federal de Minas Gerais.

<sup>2</sup> Professor Doutor. Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais. Departamento de Pós graduação em Parasitologia e Doenças Infecto-contagiosas. Área: Epidemiologia

<sup>3</sup> Professor Doutor. Instituto de Ciências Exatas da Universidade Federal de Minas Gerais. Departamento de Estatística.

**Palavras-chave:** Banco de dados; Infecção de sítio cirúrgico; Epidemiologia; Ortopedia; Estatística; Enfermagem.

## **Abstract**

Missing information in databases, especially in the health area, is a common occurrence that limits the statistical analysis of complete data. Imputation techniques of missing values were applied for surgical site infection data in orthopedic surgeries carried out in hospitals of Belo Horizonte, Minas Gerais, Brazil. The statistical analyses performed with the software of public domain R were able to compare the complete data used by Ercole (2006) with the imputed ones. Allocation procedures for missing data as described by Little & Rubin (1987) were employed according to the variable type. It was found that the allocation induced by the imputation procedure was efficient and produced similar results to those obtained with the complete data set.

**Key-Words: Database; Surgical site infection; Epidemiology; Orthopedics; Statistics; Nursing.**

## INTRODUÇÃO

Estudos que envolvem as Infecções de Sítio Cirúrgico (ISC) têm importância no contexto das Infecções Hospitalares (IH) porque estas também são responsáveis pelo aumento considerável na morbidade e mortalidade hospitalar e representam grandes custos sociais, econômicos e humanos.<sup>1-2</sup>

Uma das principais ferramentas utilizadas pelos hospitais para o armazenamento de informações coletadas no Sistema de Vigilância de Infecções Hospitalares, inclusive aquelas referentes às infecções cirúrgicas é a construção e manipulação de banco de dados.

Ao utilizar um banco de dados hospitalar em estudos epidemiológicos, deve-se considerar limitações importantes, como o fato de se trabalhar com variáveis pré-determinadas e coletadas, além do problema das perdas de informações. A ocorrência de dados faltantes (*missing data*) é uma limitação em pesquisas, especialmente na área da Saúde.<sup>3-8</sup>

O grau de precisão limitada das informações geradas a partir de banco de dados deve ser considerada porque grandes perdas de informações, especialmente para características estudadas, podem comprometer a qualidade dos resultados produzidos, gerando estimativas com erros. Estratégias para evitar e lidar com dados faltantes em variáveis importantes de estudos devem ser estabelecidas durante a fase de planejamento da pesquisa, na coleta dos dados ou na abordagem dos dados faltantes, com técnicas estatísticas adequadas e desenvolvidas para contornar o problema.<sup>7</sup>

Arnold e Kronmal (2003)<sup>9</sup> trabalharam fatores de risco para doenças cardiovasculares e enfrentaram a perda de informação em aproximadamente 150 variáveis. Ercole (2006)<sup>6</sup> trabalhou com banco de informações de ISC de 21.558 pacientes cirúrgicos ortopédicos e deparou com um percentual alto de dados perdidos

em grande parte das variáveis (até 40% em algumas variáveis), o que restringiu a análise a apenas 8.236 pacientes que apresentavam dados completos. Nunes et al<sup>10</sup> trabalharam com um conjunto de dados de 470 pacientes cirúrgicos em que havia até 20% de dados faltantes em algumas das variáveis estudadas.

Em situações com dados faltantes é comum restringir a análise aos sujeitos com dados completos nas variáveis envolvidas. Porém, as estimativas obtidas com tais análises podem levar ao viés de seleção, especialmente se os indivíduos incluídos na análise são sistematicamente diferentes daqueles que foram excluídos em termos de uma ou mais variáveis ou seja se a perda é considerada diferencial.<sup>5,10</sup>

Não há dúvidas sobre a importância de se utilizar dados secundários oriundos de banco de dados hospitalares em estudos epidemiológicos, mesmo com aqueles onde há perdas de informações, por tratar-se de uma realidade encontrada nas instituições hospitalares. Há de se ter meios de trabalhar esses dados faltantes e com isso possibilitar uma aproximação de resultados fidedignos e compatíveis com a realidade.<sup>13</sup> Assim, de que maneira essas perdas de informações podem ser trabalhadas para produzirem resultados consistentes e acurados?

Desde os anos 80 surgiram técnicas estatísticas envolvendo a substituição dos dados faltantes por estimativas de valores plausíveis a serem “imputados” aos dados faltantes. O desenvolvimento de métodos estatísticos direcionados a solucionar problemas de dados faltantes tem sido uma área de pesquisa bastante ativa nas últimas décadas.<sup>7, 11-17</sup>

Na literatura estatística esta técnica denomina-se imputação de dados faltantes e seu uso vem sendo generalizado e estendido a outras áreas. Essa técnica têm por objetivo completar os bancos de dados e possibilitar a análise com todos os indivíduos do estudo.

Assim, a proposta deste estudo é aplicar técnicas de imputação ao conjunto de dados faltantes excluídos no trabalho de Ercole (2006)<sup>6</sup> na tentativa de responder se a análise estatística após a imputação fornecerá resultados semelhantes aos obtidos no estudo anterior, quando foi restringida a análise exclusivamente às informações completas no banco de dados de cirurgias ortopédicas.

## MÉTODOS

### Fonte de dados

Foi utilizado o banco de dados do programa de controle de infecção hospitalar gerado pelo Sistema Automatizado de Controles de Infecções Hospitalares – SACIH<sup>18</sup> de pacientes atendidos em quatro hospitais de cuidado terciário, localizados em regiões distintas da cidade de Belo Horizonte, Minas Gerais, Brasil.

O banco em estudo faz parte de uma coorte histórica composta por 21.558 informações (banco original) de pacientes cirúrgicos ortopédicos (08 anos de estudo) para avaliação do risco de ISC e definição de um modelo de predição de risco de aquisição de infecção.<sup>6</sup> Entretanto, no estudo inicial, informações da maioria das variáveis não foram preenchidas nos protocolos dos pacientes cirúrgicos e por isso foram considerados dados “*missing*”. Apenas os pacientes que tinham informações para todas as características de interesse compuseram o banco completo (8.236 observações)<sup>6</sup>. Estes pacientes fizeram parte da coorte avaliada. Para essa análise a variável resposta, ISC, foi avaliada em relação às covariáveis de interesse como hospital, tipo de procedimento cirúrgico, ASA (condição clínica do paciente), potencial de contaminação da ferida cirúrgica, Índice de Risco de Infecção Cirúrgica NNIS, número de profissionais participantes da cirurgia, uso de antibiótico profilático, ocorrência de trauma, caráter emergencial da cirurgia, uso de anestesia geral, uso de prótese, tempo de cirurgia e tempo de internação pré-operatório.<sup>6</sup>

A modelagem inicial foi realizada a partir das variáveis selecionadas na análise univariada (associação estatística com a ISC com valor  $p \leq 0,20$ ) e com aquelas variáveis com significado biológico reconhecido na literatura, de forma a permitir a construção do melhor modelo alternativo dentre os vários modelos preditores construídos.<sup>6</sup> Para a análise univariada foi utilizado o teste da razão de verossimilhança, os coeficientes do

modelo de regressão logística, o *odds ratio* (OR) e valor- $p < 0,05$ . O modelo alternativo final escolhido usando o banco completo englobou as seguintes variáveis: ASA, uso de prótese, potencial de contaminação da ferida cirúrgica, hospital, número de profissionais participantes da cirurgia, uso de anestesia geral e duração da cirurgia.<sup>6</sup>

No atual estudo, foram seguidos os mesmos passos metodológicos utilizados por Ercole (2006)<sup>6</sup> em estudo anterior para a análise dos dados e alcance dos resultados pretendidos.

### **Tratamento e análise dos dados**

No presente estudo foi realizada a comparação entre o banco completo utilizado por Ercole (2006)<sup>6</sup> e o banco imputado, além de comparação dos modelos preditores de infecção, principalmente o modelo alternativo final.<sup>6</sup>

Optou-se por realizar a imputação apenas para os casos sem informação em uma característica (3.610 casos - 16,7%), duas características (2.387 casos - 11,1%) e, três características (2.196 casos - 10,2%), resultando em 8.193 casos. A decisão de restringir a imputação aos casos com perdas em até três características foi tomada para evitar que a variância dos estimadores fosse muito grande. É importante ressaltar que para fazer imputação é preciso assumir que os dados podem depender dos valores observados, mas não dependem dos dados não observados, “*missing at random*”.<sup>7</sup>

Foram realizadas imputações simples e múltiplas.

#### *Métodos de imputação*

##### *Imputação simples*

Obteve-se a amostra imputada substituindo as caselas sem informação pelo valor de maior frequência ou da média, dependendo da covariável.

##### *Imputação simples via regressão*



Neste método foram considerados os valores das demais características para estimar o valor da casela vazia. Isto foi feito baseado em modelos de regressão logístico (binária), multinomial (três categorias ou mais) ou linear (quantitativa) em dois passos.

O banco completo foi utilizado como fonte de informação para as imputações simples. O primeiro passo para fazer a imputação simples foi selecionar uma variável (aquela a ser imputada) como resposta e realizar ajustes de regressão considerando como covariáveis todas as outras características disponíveis no banco completo. Estes ajustes foram feitos considerando como características preditoras aquelas com valores-p inferiores a 0,05 no modelo cheio em todos os níveis.

As rotinas para a imputação simples foram realizadas utilizando o *software* R<sup>19</sup>, de domínio público, e as conclusões extraídas dos resultados foram obtidas considerando nível de significância igual a 5%.

#### *Imputação Múltipla*

A Imputação Múltipla – IM<sup>13, 16-17</sup> é uma técnica para analisar bancos de dados onde algumas entradas são faltantes (missing). A aplicação da técnica de imputação múltipla requer três passos: imputação, análise e agrupamento. Foi utilizado o pacote Multivariate Imputation by Chained Equations – MICE, do software de domínio público R<sup>20</sup> para fazer a imputação múltipla no banco de dados sobre ISC em cirurgias ortopédicas ([www.multiple-imputation.com](http://www.multiple-imputation.com)).

O Pacote MICE<sup>20</sup> gera imputações múltiplas para dados multivariados através do método Gibbs Sampling. O algoritmo imputa dados em uma coluna incompleta (coluna alvo) baseando-se em outras colunas do banco de dados. Cada coluna incompleta deve ser considerada como uma coluna-alvo e tem seus próprios preditores. Como padrão, o conjunto de preditores consiste de todas as outras colunas do banco de dados. Os métodos de imputação múltipla utilizados nessa análise foram PMM - Predictive Mean Matching (Numeric) e Sample - Random Sample from the observed values (Any).

### *Análise univariada da ISC dos bancos completo e incompleto*

As comparações das variáveis categóricas e quantitativas do banco completo e do imputado em relação à ISC foram realizadas a partir de tabelas de contingência com aplicação do teste Qui-Quadrado (correção de Yates) para comparação de proporções quando existiam apenas duas categorias em cada variável. Para o caso de variáveis com mais de duas categorias foi utilizado o teste Qui-Quadrado de Pearson<sup>21</sup>.

Para a comparação entre a ISC e as covariáveis quantitativas idade, tempo de cirurgia e tempo de internação pré-operatório foi utilizado o teste t-student quando as suposições usuais do modelo (normalidade - teste de Shapiro-Wilk e homocedasticidade – teste de Levene)<sup>21</sup> foram atendidas. Caso contrário foi utilizado o teste de Mann-Whitney.

### *Modelo de regressão multivariado para o banco imputado*

Para os modelos com a variável resposta ISC, foram selecionadas as covariáveis com valor-p<0,25 na análise univariada. Em seguida foram introduzidas as demais características permanecendo aquelas com significância estatística (valor-p inferior a 0,05). Foram ajustados modelos multivariados para o banco imputado com o objetivo de verificar quais seriam os resultados obtidos independentemente daqueles apresentados por Ercole (2006).<sup>6</sup> A adequação dos modelos foi verificada através do teste de Hosmer-Lemeshow.<sup>21</sup> O modelo alternativo final encontrado por Ercole (2006)<sup>6</sup> foi comparado ao modelo resultante do banco imputado.

### *Comparações entre os bancos completo, imputado e original*

Para a realização das comparações entre os casos com e sem informação (*missing*) foi criada inicialmente para cada covariável (banco original), uma variável indicadora cuja função era informar se para aquele paciente (observação) havia ou não informação em determinada variável. Em seguida cada uma das variáveis indicadoras

foi comparada com a variável resposta ISC com o objetivo de analisar se a proporção de ocorrência desta característica era a mesma entre aqueles que tinham informação para a covariável analisada e aqueles que não a tinham.

Estas comparações foram realizadas a partir de tabelas de contingência sendo aplicado a elas o teste Qui-Quadrado com correção de Yates.<sup>21</sup>

#### *Teste de razão de máxima verossimilhança para comparação dos modelos*

Os coeficientes dos modelos ajustados com o banco imputado foram comparados aos coeficientes dos modelos ajustados considerando o banco completo através de testes de razão de máxima verossimilhança. Este teste compara o modelo estimado usando o banco completo e o imputado separadamente (modelo irrestrito), com o modelo estimado considerando o agrupamento dos dois bancos, modelo restrito (16.429 dados).

A estatística do teste de razão de verossimilhança (TRV)<sup>21</sup> é a apresentada na equação abaixo, onde  $l_r$  é a verossimilhança restrita,  $l_i$  é a verossimilhança irrestrita que sob  $H_0$  (hipótese de que os coeficientes de regressão são iguais entre os modelos) segue distribuição aproximadamente Qui-quadrado com  $p$  graus de liberdade, sendo  $p$  o número de coeficientes estimados no modelo completo.

$$TRV = -2 \log \left( \frac{l_r}{l_i} \right) \stackrel{H_0}{\sim} \chi^2_p$$

Os testes para comparação dos coeficientes foram realizados considerando tanto o banco imputado de forma simples como aquele imputado via regressão.

#### *Considerações éticas e legais do estudo*

O projeto de pesquisa (Processo ETIC 274/04) foi submetido ao Comitê de Ética e Pesquisa da Universidade Federal de Minas Gerais, sendo julgado e aprovado pelo

mesmo, uma vez que o estudo envolve dados secundários. Durante a realização deste trabalho não houve nenhum tipo de conflito de interesse entre os participantes.

## RESULTADOS

Análise descritiva das variáveis independentes do banco completo, imputado e original.

A descrição das variáveis de interesse considerando os bancos completo e imputado (até três características faltantes de forma simples e via modelos de regressão) e o original, com perdas de informações, podem ser analisados na Tabela 1.

As porcentagens de casos imputados (colunas de imputação simples e regressão) ou as porcentagens de casos sem informação (21.558 observações) são apresentadas em negrito na Tabela 1. Considere o exemplo: para a variável ASA foram imputados 782 casos, ou seja, 9,5% do total de 8.193; já para todo o banco observa-se que não há informação para 3.584 (16,6%) pacientes do total de 21.558.

Os Odds Ratio (OR) dos ajustes realizados com o banco de dados completos e aquele com dados imputados de forma simples e usando a regressão são apresentadas na Tabela 2.

A partir das variáveis iniciais do banco original foram construídos modelos denominados Modelo 0, Modelo 1, Modelo 2, Modelo 3, Modelo 4 e Modelo 5 (Tabela 2), tanto para o banco completo quanto para o banco imputado das duas formas, simples e via regressão. Ressalta-se que as variáveis que entraram no Modelo alternativo final de predição de ISC, encontrado em Ercole (2006)<sup>6</sup> coincide com as variáveis do Modelo 2 (Tabela 2).

As estimativas das OR das categorias com diferença significativa (valor- $p \leq 0,05$ ) estão em negrito e aquelas com tendência de significância estatística (valor- $p \leq 0,10$ ) apresentam-se com asteriscos. Observa-se grande diferença das OR quando comparada o escore 3 do Índice de Risco de Infecção Cirúrgica NNIS em relação ao escore 0 nos modelos 1, 4 e 5. Isso pode ser explicado pelo fato de esta característica conter 3.830 valores perdidos. Os demais casos não aparentam ter grandes diferenças em relação à

OR. Em particular, não se observam diferenças significativas das OR do Modelo 2 em relação aos bancos imputado e completo.

Teste de razão de máxima verossimilhança para comparação dos modelos

A Tabela 3 apresenta o número de coeficientes ajustados em cada um dos modelos, bem como o valor da função de log-verossimilhança do banco completo e do banco imputado de forma simples e via regressão, além dos valores da log-verossimilhança para o banco agrupado, tanto no caso de imputação simples, quanto via regressão e os valores-p para cada modelo.

Observa-se que o Modelo 2 (Tabela 3), corresponde ao modelo alternativo final encontrado por Ercole (2006)<sup>6</sup>. Neste modelo não há diferenças em relação aos coeficientes de regressão nos bancos completo e imputado, significando que se pode considerar para a análise estatística, um só banco contendo os dados completos e imputados. Apenas o Modelo 0, que contém somente a variável Índice de Risco de Infecção Cirúrgica NNIS, apresentou diferença significativa entre os coeficientes dos dois bancos. Este resultado já era esperado uma vez que este modelo não é um bom preditor de ISC.

De uma maneira geral, os achados deste estudo indicam a possibilidade de utilização dos resultados do banco agrupado (completo + imputado).

Imputação Múltipla

O teste da razão da verossimilhança realizado para os coeficientes do Modelo 2 encontrado utilizando a imputação múltipla mostra resultado semelhante ( $p < 0,05$ ) ao observado com a imputação simples.

A Tabela 4 apresenta os coeficientes estimados para o Modelo 2 usando imputação múltipla para o banco imputado (8.193). Este resultado reforça o fato da

possibilidade de utilização dos dois bancos agrupados, ou seja, a imputação não afetou a estimação dos modelos.

## DISCUSSÃO

A ocorrência de dados faltantes é comum em investigações científicas, especialmente as pesquisas na área da saúde que trabalham com banco de dados. Como os dados faltantes podem causar problemas nas estimativas (viés) é necessário considerá-los nas análises. A imputação de dados faltantes é um método estatístico que substitui os dados sem informação por estimativas de valores plausíveis, direcionado a solucionar este tipo de problema.<sup>3-4, 6-7</sup>

Apesar de ser um método cuja aplicação esteja em expansão desde o início da década de 90, pode-se afirmar que são poucos os trabalhos na área da saúde<sup>7</sup>, especialmente, os que tratam de banco de dados sobre ISC e que se encontram publicados. Assim, trabalhos nessa área devem ser desenvolvidos e divulgados antes que se possam utilizar os métodos de imputação rotineiramente e com confiança.

Vale ressaltar que o objetivo deste artigo foi utilizar as técnicas de imputação simples e múltipla para contornar o problema da grande quantidade de dados faltantes no banco de ISC em cirurgias ortopédicas e comparar os resultados das análises, anteriormente restrita ao banco de dados com apenas as informações completas (8.236)<sup>6</sup>, com aquelas constantes no banco de 8.193 dados faltantes imputados.

Assim, foi considerada adequada a utilização das técnicas de imputação para contornar o problema da grande quantidade de dados faltantes (16.429) na maioria das variáveis do banco original (21.558) de ISC com a intenção de obter melhores resultados. Restringir a análise ao conjunto de dados com observações completas pode levar a resultados e conclusões incorretas.<sup>11, 14</sup>

Uma vantagem deste estudo em relação aos demais, onde são utilizadas técnicas de imputação, é que o banco completo é grande, com 8.236 pacientes. Com uma amostra deste tamanho o erro das estimativas de grande parte dos parâmetros é



pequeno.<sup>21</sup> Desta forma, as comparações estabelecidas entre as estimativas geradas com o banco completo e com o imputado são confiáveis em relação àquelas feitas em bancos com número pequeno de informações, como pode ser visto em trabalho de simulação para dados epidemiológicos realizado por Nunes e colaboradores.<sup>7</sup>

Do total de 21.558 informações de pacientes cirúrgicos ortopédicos contidos no banco original, 5.129 tinham mais do que três perdas por paciente em cada variável (mais de três dados faltantes). As 16.429 informações de pacientes cirúrgicos ortopédicos tinham entre 0 e 3 perdas por paciente, sendo que 8.236 tinham todas as informações completas e 8.193 tinham entre 1 e 3 dados faltantes. Neste estudo não foram consideradas as cirurgias em que faltaram informações em mais de três variáveis do banco. Isso foi feito porque o erro total de estimação dos parâmetros do modelo resultante poderia inviabilizar as análises estatísticas.<sup>4,8,10,12</sup> Portanto, as técnicas de imputação foram realizadas no banco com 8.193 informações faltantes.

É importante ressaltar que nas 5.129 cirurgias que não entraram na análise atual há diferenças estatisticamente significativas no que se refere ao risco de infecção cirúrgica. Nesse conjunto de dados (5.129) a taxa de infecção foi de 2,8%, quando comparado à taxa de infecção de 1,7% do conjunto de dados (16.429) que tiveram entre 0 e 3 dados faltantes.

A princípio este resultado indicaria que as perdas seriam diferenciais e as estimativas dos parâmetros do modelo poderiam estar viesadas.<sup>4,8,10</sup> Mas, a análise descritiva dos dados com informações faltantes mostra que a maioria das perdas (4.427) vem de um único hospital (hospital 0) e que, também, neste hospital ocorre a maior taxa de ISC. Assim, quando as informações de pacientes submetidos a cirurgias ortopédicas contidas em variáveis com maior perda são desconsideradas, não são consideradas também aquelas com maior taxa de ISC. Portanto, pode-se inferir que as estimativas

encontradas não são viciadas, mas não podem ser consideradas válidas para o Serviço de Controle de Infecção Hospitalar deste hospital uma vez que podem existir problemas no gerenciamento das informações sobre os pacientes cirúrgicos ortopédicos.

Os coeficientes de regressão do Modelo Alternativo Final, encontrado por Ercole (2006)<sup>6</sup>, estimados usando o banco completo (8.236) e o encontrado neste estudo obtido com o banco imputado (8.193) não apresentaram diferenças significativas, de acordo com o teste de razão de verossimilhança. Os resultados estatísticos são compatíveis com os obtidos em estudo sobre predição de risco de infecção ortopédica<sup>6</sup>. Os dois bancos podem ser agrupados em um único (16.429 dados), uma vez que os resultados das estimativas obtidas para os coeficientes do modelo de regressão são iguais.

Um resultado importante deste estudo é que os coeficientes dos modelos de regressão estimados usando o banco completo e o banco imputado não apresentaram diferenças significativas, de acordo com o teste de razão de verossimilhança. Isto significa que os dois bancos podem ser agrupados e chegar às mesmas estimativas para os coeficientes do modelo de regressão. Assim, o aumento no tamanho do conjunto de dados torna as análises mais poderosas.

## **CONCLUSÃO**

Concluiu-se que realizar imputações no banco com dados faltantes não alterou as conclusões das análises estatísticas, fato este que pode indicar positivamente o uso desta estratégia para analisar bancos com grandes perdas de informações. Apesar dos resultados encontrados terem sido, especificamente, para dados relativos à ISC de pacientes ortopédicos, acredita-se que a imputação gerou resultados confiáveis, podendo ser uma solução para bancos de dados de infecção e também para outros tipos de bancos com grandes perdas de informações que tenham características semelhantes.

Acredita-se ser necessária a realização de mais estudos na área da saúde que indiquem as vantagens do uso dos métodos de imputação para tratar o problema dos dados faltantes em diferentes tipos de bancos de dados.

## REFERÊNCIAS

1. Proveda VB, Galvão CM, Hayashida M. Análise dos fatores de risco relacionados à incidência de infecção do sítio cirúrgico em gastrocirurgias. *Rev Esc Enferm USP* 2003; 37(1): 81-89.
2. Perl TM. Prevention of *Staphylococcus aureus* infections among surgical patients: beyond traditional perioperative prophylaxis. *Surgery* 2003; 134 (Suppl 5): 10-17.
3. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc* 1996; 91(434): 473-89.
4. Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons; 1987. 258p.
5. Machado CJ. Early infant morbidity and infant mortality in the city of São Paulo, Brazil: a probabilistic record linkage approach. [Tese de Doutorado em Filosofia]. Baltimore, Maryland: Johns Hopkins University; 2002.
6. Ercole FF. Avaliação da aplicabilidade do índice de risco de infecção cirúrgica do sistema NNIS em pacientes submetidos a procedimentos ortopédicos: um estudo multicêntrico em hospitais de Belo Horizonte, Minas Gerais. [Tese de Doutorado]. Belo Horizonte: – Departamento de Parasitologia, Universidade Federal de Minas Gerais; 2006.
7. Nunes LN. Métodos de imputação de dados aplicados na área da saúde. [Tese de Doutorado]. Porto Alegre: Faculdade de Medicina, Universidade Federal do Rio Grande do Sul; 2007.
8. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol* 2006; 59:1102-9.

9. Arnold AM, Kronmal RA. Multiple imputation of baseline data in the Cardiovascular Health Study. *American Journal of Epidemiology* 2003; 157(1): 74-84.
10. Nunes LN, Kluck MM, Fachel JMG. Uso da imputação múltipla de dados faltantes: uma simulação utilizando dados epidemiológicos. *Cadernos de Saúde Pública* [online] 2009; 25(2): 268-278.
11. Szklo M, Javier Nieto F. *Epidemiology: beyond the basics*. Gaithersburg/Maryland: AN Aspen Publication. 2000. 493p.
12. Harel O, Zhou XH. Multiple imputation: review of theory, implementation and software. *Statistics in Medicine* 2007; 26(16): 3057-77.
13. Little RJA. Regression with Missing Xs: a review. *Journal of the American Statistical Association* 1992; 87(420): 1227-37.
14. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Research* 1999; 8:3-15.
15. Zhang P. Multiple imputation: theory and method. *International Statistical Review* 2003; 71(3): 581-92.
16. Kenward, MG, Carpenter J. Multiple imputation: current perspectives. *Stat Methods Med Res* 2007; 16:199-218.
17. Little RJA, Rubin DB. *Statistical analysis with missing data*. 1ª ed. EUA: John Wiley & Sons, 1987. 278p.
18. Starling CEF, Pinheiro SMC, Couto BRGM. *Vigilância epidemiológica das infecções hospitalares na prática diária: ensaios*. Belo Horizonte: Cuatiara, 1993. 488p.
19. Casella G, Berger RL. *Statistical Inference*. 2ª ed. Belmont, CA: Duxbury; 2002. 660p.
20. van Buuren S, Oudshoorn CGM. *Multivariate imputation by chained equations*. MICE V1.0 User's manual. Leiden: TNO Preventie en Gezondheid; 2000. 39p.

21. Hosmer JR, David W, Lemeshow S. Applied logistic regression. 1<sup>th</sup>ed. New York: John Wiley & Sons; 1989. 307p.

### **Agradecimentos**

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG pelo apoio e subsídio financeiro desta pesquisa e ao Dr. Carlos Ernesto Ferreira Starling, Médico Infectologista e Coordenador dos Serviços de Infecção Hospitalar dos hospitais do estudo, por ter cedido o banco de dados analisado.

Tabela 1

Descrição das variáveis de interesse considerando o banco completo, o imputado pela forma simples e via regressão e banco original. Belo Horizonte, MG – 2009.

Variáveis	Frequência						
	Banco Completo		Banco Imputado - 8.193				Banco Original
	8.236		Imputação Simples		Imputação Regressão		21.558
	n	%	n	%	n	%	Total n
Hospital							
0	2.752	33,4	4.295	52,4	4.295	52,4	11.474
1	4.000	48,6	2.335	28,5	2.335	28,5	6.525
2	563	6,8	497	6,1	497	6,1	1.262
3	921	11,2	1.066	13	1.066	13	2.297
ASA – condição clínica do paciente							
ASA I	5.021	60,9	5.929	72,4	5.763	70,3	12.131
ASA II	2.378	28,9	1.740	21,2	1.906	23,4	4.413
ASA III	763	9,3	475	5,8	475	5,8	1.297
ASA IV	74	0,9	49	0,6	49	0,6	133
<b>Sem informação/imputados</b>	-	-	<b>782</b>	<b>9,5</b>	<b>782</b>	<b>9,5</b>	<b>3.584</b>
Tipo de procedimento cirúrgico							
US	1.232	15	1.504	18,3	1.504	18,3	3.628
FX	2.905	35,3	2.582	31,5	2.582	31,5	7.274
OMS	3.429	41,6	3.605	44	3.605	44	9.371
PROS – J	131	1,6	61	0,7	61	0,7	204
PROS – O	91	1,1	40	0,5	40	0,5	141
PROS – Q	448	5,4	401	0,9	401	0,9	940
Potencial de contaminação da ferida cirúrgica							
Limpa	7.176	87,2	6.927	84,6	6.925	84,5	14.271
Potencialmente contaminada	560	6,8	720	8,8	720	8,8	1.355
Contaminada	184	2,2	206	2,5	206	2,5	410
Infectada	316	3,8	340	4,1	342	4,2	698
<b>Sem informação/imputados</b>	-	-	<b>491</b>	<b>6</b>	<b>491</b>	<b>6</b>	<b>4.824</b>
Índice de Risco de Infecção Cirúrgica NNIS							
Escore 0	5.263	63,9	6.443	78,6	6.138	74,9	7.995
Escore 1	2.523	30,6	1.487	18,2	1.792	21,9	4.046
Escore 2	431	5,2	252	3,1	252	3,1	685
Escore 3	19	0,2	11	0,1	11	0,1	30
<b>Sem informação/imputados</b>	-	-	<b>3.830</b>	<b>46,8</b>	<b>3.830</b>	<b>46,8</b>	<b>8.802</b>
Uso de antibiótico profilático							

continuação Tabela 1

Não	3.602	43,7	1.997	24,4	4.000	48,8	5.773
Cefazolina	4.472	54,3	6.100	74,5	4.097	50	7.043
Clindamicina + outros	162	2	96	1,1	96	1,2	266
<b>Sem informação/imputados</b>	-	-	<b>3.867</b>	<b>47,2</b>	<b>3.867</b>	<b>47,2</b>	<b>8.476</b>
Tempo de cirurgia							
< 120 minutos	4.614	56	5.576	68,1	5.324	65	8.493
≥ 120 minutos	3.622	44	2.617	31,9	2.869	35	6.666
<b>Sem informação/imputados</b>	-	-	<b>2.832</b>	<b>34,6</b>	<b>2.832</b>	<b>34,6</b>	<b>6.399</b>
Número de profissionais participantes da cirurgia							
1 a 4 profissionais	3.799	46,1	5.275	64,4	5.273	64,4	12.240
5 a 8 profissionais	4.367	53	2.867	35	2.869	35	7.923
9 a 16 profissionais	70	0,9	51	0,6	51	0,6	146
<b>Sem informação/imputados</b>	-	-	<b>476</b>	<b>5,8</b>	<b>476</b>	<b>5,8</b>	<b>1.249</b>
Uso de anestesia geral							
Sim	3.778	45,9	3.468	42,3	3.716	45,4	8.740
Não	4.458	54,1	4.725	57,7	4.477	54,6	9.115
<b>Sem informação/imputados</b>	-	-	<b>1.184</b>	<b>14,5</b>	<b>1.184</b>	<b>14,5</b>	<b>3.703</b>
Caráter emergencial da cirurgia							
Sim	1.138	13,8	867	10,6	880	10,7	2.101
Não	7.098	86,2	7.326	89,4	7.313	89,3	14.670
<b>Sem informação/imputados</b>	-	-	<b>389</b>	<b>4,8</b>	<b>389</b>	<b>4,8</b>	<b>4.787</b>
Trauma							
Sim	3.392	41,2	2.480	30,3	2.585	31,5	6.045
Não	4.844	58,8	5.713	69,7	5.608	68,5	10.754
<b>Sem informação/imputados</b>	-	-	<b>318</b>	<b>3,9</b>	<b>318</b>	<b>3,9</b>	<b>4.759</b>
Uso de prótese							
Sim	4.004	48,6	3.292	40,2	3.722	45,4	7.506
Não	4.232	51,4	4.901	59,8	4.471	54,6	8.680
<b>Sem informação/imputados</b>	-	-	<b>770</b>	<b>9,4</b>	<b>770</b>	<b>9,4</b>	<b>5.372</b>
Infecção do sítio cirúrgico (ISC)							
Sim	116	1,4	165	2	165	2	431
Não	8.120	98,6	8.028	98	8.028	98	21081
<b>Sem informação/imputados</b>	-	-	<b>33</b>	<b>0,4</b>	<b>33</b>	<b>0,4</b>	<b>46</b>



**Tabela 2**

**Odds Ratio dos modelos ajustados considerando o banco completo e o imputado.**

**Belo Horizonte, MG - 2009.**

Variáveis	Modelo 0			Modelo 1			Modelo 2			Modelo 3			Modelo 4			Modelo 5		
	C	S	R	C	S	R	C	S	R	C	S	R	C	S	R	C	S	R
<b>Hospital</b>																		
0	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
1	-	-	-	-	-	-	0,2	0,3	0,3	0,2	0,4	0,3	0,2	0,3	0,3	0,2	0,3	0,2
2	-	-	-	-	-	-	0,5*	0,4	0,4	0,7	0,4	0,4	0,7	0,4	0,4	0,5*	0,4	0,4
3	-	-	-	-	-	-	0,3	0,4	0,4	0,4	0,4	0,4	0,4	0,4	0,4	0,3	0,3	0,3
<b>Asa-Condição Clínica do Paciente</b>																		
ASA I	-	-	-	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-
ASA II	-	-	-	1,4	1,4*	1,5	1,9	1,6	1,7	2	1,6	1,7	-	-	-	-	-	-
ASA III	-	-	-	1,9	1,6	1,6	3	2	2,2	3,1	2	2,1	-	-	-	-	-	-
ASA IV	-	-	-	2,6	3,0*	3,0*	4,8	4	4,3	4,7	4	4,3	-	-	-	-	-	-
<b>Tempo cirurgia</b>																		
≥120 min	-	-	-	1,2	1,4*	1,3*	1,7	1,6	1,4*	-	-	-	-	-	-	-	-	-
<120 min	-	-	-	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-
<b>Potencial de contaminação da ferida cirúrgica</b>																		
Limpa	-	-	-	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-
Potencialmente contaminada	-	-	-	2	2,4	2,5	2,5	2,1	2,1	2,3	2	2,1	-	-	-	-	-	-
Contaminada	-	-	-	3,4	1,6*	1,6	4,8	2,2*	2,3*	4,3	2,2*	2,3*	-	-	-	-	-	-
Infectada	-	-	-	3	2,7	3,1	3,9	2,9	3,4	3,6	2,8	3,4	-	-	-	-	-	-
<b>Índice de Risco de Infecção Cirúrgica NNIS</b>																		
Escore 0	1	1	1	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1
Escore 1	1,8	1,4	1,3	-	-	-	-	-	-	-	-	-	2,7	1,5	1,6	2,6	1,5	1,6
Escore 2	2,7	2,5	2,5	-	-	-	-	-	-	-	-	-	5,1	3,8	4,1	4,3	3,5	3,8
Escore 3	5,2	20,6	20,7	-	-	-	-	-	-	-	-	-	11,3	44,9	52,8	10,2	41,5	48,6
<b>N. de profissionais na sala de cirurgia</b>																		
1 a 4	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
5 a 8	-	-	-	-	-	-	1,1	1,3	1,2	1,1	1,3	1,3	1,1	1,3	1,3	1,1	1,3	1,2
9 a 16	-	-	-	-	-	-	5,3	1,9	1,7	5,9	2,1	2,1	5,5	2,2	2,1	5,4	2,2	2,1
<b>Anestesia geral</b>																		
Sim	-	-	-	-	-	-	0,7*	1,1	1,1	-	-	-	-	-	-	-	-	-
Não	-	-	-	-	-	-	1	1	1	-	-	-	-	-	-	-	-	-
<b>Uso prótese</b>																		
Sim	-	-	-	-	-	-	1,9	2	2,1	2,1	2,2	2,2	1,6	2,1	2,1	1,6	2	2
Não	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
<b>Idade (contínua)</b>																		
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,01	1,01	1,01

Nota: -- não está presente no modelo; \* valor-p<0,10 na análise multivariada (tendência de significância estatística); C: Banco completo; I: Banco Imputado; S: Imputação simples; R: Imputação regressão

**Tabela 3**

Valor da razão de máxima verossimilhança para cada modelo. Belo Horizonte, MG - 2009.

Modelos	nº de coeficientes	Banco Completo 8.236	Banco Imputado		Banco Completo + Imputado		Valor-p
			8.193		16.429		
			Imputação Simples	Imputação Regressão	Imputação Simples	Imputação Regressão	
0	4	-602,60	-798,1	-798,19	-1408,79	-1408,55	0,003
1	8	-595,05	-790,88	-788,09	-1393,24	-1390,09	0,067
2	15	-565,30	-763,99	-761,28	-1336,86	-1332,71	0,441
3	13	-569,41	-767,20	-762,88	-1342,85	-1337,07	0,488
4	10	-577,87	-770,53	-769,35	-1357,17	-1354,54	0,063
5	11	-575,73	-771,67	-768,41	-1353,52	-1351,12	0,121

**Tabela 4**  
**Coefficientes estimados através da imputação múltipla para**  
**o Modelo 2 de predição de risco de ISC. Belo Horizonte, MG - 2009.**

Variável	Modelo 2	Imputação Múltipla	
		Sample	pmm
<b>Constante</b>	<b>-4,406</b>	<b>-3,892</b>	<b>-3,89</b>
<b>Asa – Condição Clínica do Paciente</b>			
asa_2	0,633	0,337	0,455
asa_3	1,111	0,786	0,978
asa_4	1,576	1,221	1,483
<b>Hospital</b>			
Hospit_1	-1,758	-1,431	-1,454
Hospit_2	-0,655	-0,72	-0,722
Hospit_3	-1,363	-0,747	-0,742
<b>Tempo de cirurgia</b>			
Tempo_1	0,503	1,002	1,002
<b>Potencial de contaminação da ferida cirúrgica</b>			
Pot. cont_2	0,904	0,794	1,424
Pot. cont_3	1,564	0,747	1,265
Pot. cont_4	1,347	-	-
<b>Uso anestesia geral</b>			
Sim_1	0,383	0,862	1,035
<b>Uso de prótese</b>			
Sim_1	0,628	0,517	0,439
<b>N. de profissionais na sala de cirurgia</b>			
N. de profissionais_1	0,065	0,282	0,191
N. de profissionais_2	1,167	0,798	0,597