Imputação múltipla e análise de casos completos em modelos de regressão logística: uma avaliação prática do impacto das perdas em covariáveis

Multiple imputation and complete case analysis in logistic regression models: a practical assessment of the impact of incomplete covariate data

Vitor Passos Camargos [1,2]
Cibele Comini César [3]
Waleska Teixeira Caiaffa [1,2]
Cesar Coelho Xavier [1,2]
Fernando Augusto Proietti [1,2]

[1] Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
[2] Observatório de Saúde Urbana, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
[3] Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

Correspondência
V. P. Camargos
Programa de Pós-graduação em Saúde Pública, Faculdade de Medicina, Universidade Federal de Minas Gerais.
Av. Prof. Alfredo Balena 190, sala 625, Belo Horizonte, MG 30130-100, Brasil.
vitorcamargos@gmail.com

Abstract

Researchers in the health field often deal with the problem of incomplete databases. Complete Case Analysis (CCA), which restricts the analysis to subjects with complete data, reduces the sample size and may result in biased estimates. Based on statistical grounds, Multiple Imputation (MI) uses all collected data and is recommended as an alternative to CCA. Data from the study Saúde em Beagá, attended by 4,048 adults from two of nine health districts in the city of Belo Horizonte, Minas Gerais State, Brazil, in 2008-2009, were used to evaluate CCA and different MI approaches in the context of logistic models with incomplete covariate data. Peculiarities in some variables in this study allowed analyzing a situation in which the missing covariate data are recovered and thus the results before and after recovery are compared. Based on the analysis, even the more simplistic MI approach performed better than CCA, since it was closer to the post-recovery results.

Methods; Statistical Analysis; Logistic Models; Body Mass Index

Introdução

Um problema frequente nos inquéritos populacionais de saúde é a incompletude das bases de dados. Como passo anterior às análises, a maioria dos *softwares* estatísticos adota como procedimento padrão a exclusão dos indivíduos com uma ou mais informações ausentes. Esse método, conhecido como Análise de Casos Completos (ACC), pode produzir estimativas viciadas quando suposições necessárias à sua aplicação são violadas [1]. Mesmo quando essas suposições são válidas, a perda de poder devido à exclusão dos indivíduos, que na prática reduz o tamanho da amostra, é inevitável.

Destacam-se como alternativa à ACC, com ampla literatura de estudos comparativos [1,2], os métodos de Ponderação de Casos Completos, Imputação Simples, Imputação Múltipla (IM) e Máxima Verossimilhança (MV). Esses estudos apontam os métodos de MV e IM como referenciais para a análise de bases incompletas, já que ambos além de utilizar todas as informações coletadas devem produzir resultados válidos sob condições menos restritas que a ACC. A Análise de Casos Completos usualmente requer que os indivíduos não excluídos das análises possam ser vistos como uma amostra aleatória de todos os participantes do estudo. Discute-se ainda vantagens da IM sobre o método de MV em relação à praticidade de aplicação e disponibilidade, uma vez que apenas o primeiro encontra-se im-

plementado na maioria dos *softwares* de análise tradicionais [3].

A ausência de informações é recorrente em temas sensíveis (e.g. renda) e também em indicadores compostos, nos quais as perdas nas variáveis de composição tendem a se somar. O índice de massa corporal (IMC), determinado pela divisão do peso (em quilos) do indivíduo pelo quadrado de sua altura (em metros), é um indicador de grande relevância na saúde pública. Utilizado nas pesquisas como marcador de sobrepeso e obesidade, o IMC é preditor de diversas doenças não transmissíveis como diabetes, doenças cardiovasculares e câncer [4]. Ele pode ser obtido por medidas autorreferidas de peso e altura, que aqui será denominado IMC_referido, ou, por medidas diretamente aferidas, denominado IMC_direto. Motivos diversos como constrangimento, desconhecimento ou receio de relatar informações incorretas podem explicar a ausência das medidas autorreferidas de peso e altura, tornando o IMC_referido especialmente vulnerável às perdas. Espera-se, assim, que as perdas nessa variável não ocorram de forma completamente aleatória. Aspectos práticos das análises de uma base de dados na qual o IMC_referido apresentou elevado percentual de perdas serão discutidos aqui.

A validade dos resultados dos diferentes métodos de análise de dados incompletos depende de suposições sobre os mecanismos ou fatores associados às perdas. Rubin [5] introduziu três mecanismos teóricos gerais que regulam a ocorrência das perdas e que são extensamente utilizados na literatura: Perda Completamente Aleatória (PCA), Perda Aleatória (PA) e Perda Não Aleatória (PNA). O termo mecanismo é utilizado como sinônimo da função de distribuição dos dados ausentes, que define a probabilidade de cada valor ser ou não observado, e os fatores associados a essa probabilidade. A Tabela 1 introduz esses mecanismos e apresenta exemplos adaptados de Sterne et al. [6], aqui relacionados à ausência de valores do IMC.

Na prática, embora seja plausível para questões como uso de drogas ou renda, a PNA é uma suposição que não pode ser verificada apenas com os dados observados. O impacto da PNA nas análises produzidas pelos métodos de ACC, IM e outros tem sido majoritariamente avaliado por estudos de simulação [2,7,8]. Apesar da importância desses estudos, Graham [9] argumenta a favor dessa mesma avaliação em situações práticas. Isso seria possível pela recuperação parcial dos dados usando-se uma amostra aleatória daqueles inicialmente ausentes [10].

Neste estudo, dados de um inquérito de saúde domiciliar, denominado *Saúde em Beagá* (SBH), são analisados e validados por uma abordagem similar à recuperação de dados ausentes, com dois objetivos principais: (a) checar a suposição de PNA na variável IMC_referido; (b) comparar as estimativas dos coeficientes de modelos de regressão logística segundo as abordagens de ACC e IM, na situação em que a perda ocorre predominantemente em uma única covariável (IMC), com as estimativas obtidas na base de dados sem perdas. A seguir são revisados

Tabela 1

Tipos de perda em uma variável de interesse (demais variáveis são completamente observadas).

Tipo de perda	Exemplo
Perda Completamente Aleatória	A probabilidade de perda não depende das variáveis presentes no estudo, sendo, portanto, constante para todos os indivíduos. Esse mecanismo dificilmente se verifica na prática, mas poderia ocorrer, por exemplo, para valores ausentes do IMC resultantes de erros de digitação durante a entrada dos questionários.
Perda Aleatória	A probabilidade de perda está relacionada com outras variáveis do estudo, mas não com a variável de interesse, ou seja, a probabilidade de perda está relacionada com um subconjunto conhecido dos dados. Por exemplo, valores ausentes do IMC referido podem ser maiores do que os observados se pessoas fisicamente inativas tiverem maior proporção de perda devido ao acesso menos regular a seu peso.
Perda Não Aleatória	Ocorre quando a probabilidade de perda está relacionada com os valores da própria variável de interesse, que não foram observados – assim, essa relação é desconhecida. Por exemplo, pessoas com valores extremos do IMC podem se sentir menos à vontade para relatar seu peso do que as demais.

IMC: índice de massa corporal.

os conceitos básicos do método de IM e ACC com enfoque nos modelos de regressão logística com covariáveis incompletas. Na seção Métodos apresentamos aspectos do inquérito e da abordagem de validação das análises, que buscaram aproximar a situação de recuperação dos dados ausentes do IMC_referido. Finalmente, são descritas as variáveis utilizadas nos modelos de regressão logística e as diferentes abordagens de imputação avaliadas.

Métodos

Imputação múltipla

Proposta por Rubin [11] em 1978, a IM é um método para lidar com o problema da análise de bases incompletas. A IM está disponível nos principais *softwares* estatísticos comerciais e gratuitos. Horton & Kleinman [3] apresentam uma boa revisão dessas implementações. O método de IM, se corretamente aplicado, deve produzir estimativas não viciadas mesmo sob os mecanismos de PCA e PA [1,6]. Para ilustrar o método de IM suponha uma base de dados em que valores da variável X_1 estejam ausentes e que outras variáveis sejam completamente observadas. O processo de imputação e análise dos dados pode ser dividido em quatro etapas principais:

- Passo 1 (modelo de imputação): Após selecionar as variáveis que serão utilizadas no processo de imputação busca-se, para cada dado ausente da variável X_1, valores plausíveis para o processo de imputação. A ideia básica é avaliar a distribuição dos valores observados de X_1 para indivíduos com o mesmo perfil (valores idênticos nas variáveis selecionadas) daquele com o dado ausente. Para isso, usualmente estima-se a distribuição preditiva de X_1 condicionada nas variáveis selecionadas, seguindo uma abordagem bayesiana.
- Passo 2: Cada valor ausente é substituído por M valores aleatoriamente amostrados da distribuição condicional preditiva. Esse processo produz ao final M versões completas do banco de dados.
- Passo 3 (modelo de análise): Cada banco de dados é analisado pelos métodos usuais para dados completos.
- Passo 4: Os resultados das M análises são combinados de modo a produzir estimativas que levam em conta a incerteza dos valores imputados.

Os passos 2, 3 e 4 demandam pouco trabalho adicional para o pesquisador quando comparado àquele exigido na ACC. No entanto, o primeiro passo é mais trabalhoso e decisivo para a validade dos resultados produzidos pelas análises subsequentes. Nesse, o pesquisador deve definir as variáveis que farão parte do modelo de predição, também conhecido como modelo de imputação, e o tipo de modelo (por exemplo, linear, logístico ou multinomial) que melhor se ajusta à distribuição de X_1. van Buuren et. al. [12] propõem uma estratégia geral e bem fundamentada para a seleção de variáveis: (1) incluir todas as variáveis que serão utilizadas em análises conjuntas com X_1, que é frequentemente denominado modelo de análise; (2) incluir variáveis associadas às perdas; (3) incluir variáveis preditoras de X_1; (4) excluir das etapas 2 e 3 aquelas variáveis que apresentam uma elevada proporção de perdas para aqueles em que X_1 é ausente.

Para ilustrar alguns cuidados necessários ao processo de imputação, considere o contexto no qual o modelo de análise é um modelo logístico em que X_1 (e.g. IMC) é uma covariável sujeita a perdas e que as demais covariáveis $X_{2:k}$ (e.g. idade e escolaridade) e a variável dependente Y (e.g. diagnóstico de diabetes) são completamente observadas. Considere ainda que o modelo de imputação de X_1 inclui como preditoras apenas as variáveis do modelo logístico (Y e $X_{2:k}$).

Se a perda em X_1 está associada apenas às variáveis utilizadas no modelo de imputação, as estimativas resultantes da IM são válidas (mecanismo de PA). Entretanto, se o mecanismo de PNA atua em X_1, mesmo quando consideramos indivíduos com o mesmo perfil em Y e $X_{2:k}$, as estimativas dos coeficientes dos modelos logísticos baseadas na IM podem estar viciadas. O mecanismo de PNA pode ser resultante da omissão de variáveis no modelo de imputação que estão associadas às perdas. Assim, além de incluir as variáveis presentes no modelo logístico, deve-se incluir ainda variáveis potencialmente associadas às perdas, principalmente se estas também estiverem associadas aos valores de X_1 [8].

Como regra geral, todas as variáveis (Y e $X_{2:k}$) e complexidades (estratificação das análises ou efeitos de interação) consideradas no modelo logístico devem ser incluídas no modelo de imputação. Por exemplo, a omissão de Y no modelo de imputação de X_1 pressupõe a inexistência de uma associação direta entre ambas, e assim, a *odds ratio* (OR) de X_1 estimada pela IM é viciada em direção a 1 [13]. Se modelos logísticos distintos serão avaliados em estratos da população de estudo (e.g. sexo), para que possíveis interações entre os estratos e a distribuição de X_1 sejam preservadas, o modelo de imputação também deve ser aplicado independentemente em cada estrato [2].

O modelo de imputação, assim como outros modelos de predição, envolve preocupações relacionadas à preservação de características importantes da distribuição de X_1. Se X_1 é uma

variável contínua com distribuição claramente não normal, recomenda-se a aplicação de alguma transformação (aplicação de uma função que admita uma única inversa, e.g., logarítmica, cuja inversa é a função exponencial) antes do processo de imputação, já que o modelo preditivo linear usualmente envolve a suposição de normalidade. Nesse caso, após a imputação, aplica-se a transformação inversa para que a variável retorne à sua escala original [2].

Outro aspecto importante a ser considerado na IM é a escolha do número de imputações (*M*). A escolha de um *M* pequeno pode inflacionar o intervalo de confiança das estimativas e consequentemente reduzir o poder das análises. Rubin [13] quantifica essa inflação para diferentes escolhas de *M* e frações de informação ausente (mede o quanto a variância da estimativa de um parâmetro, e.g., coeficiente de regressão, na abordagem de IM está inflacionada devido às perdas). Por exemplo, para frações de informação ausente de 20%, valores de $M = 3$ ou $M = 5$, produzem respectivamente intervalos apenas 3% e 1% mais largos do que os ideais (intervalos mínimos que seriam alcançados quando *M* tende ao infinito). O conceito de fração de informação ausente está diretamente relacionado à proporção de dados ausentes, e no caso de estimativas básicas, como por exemplo, a média de X_1, o valor da fração de informação ausente desta estimativa é igual à proporção de dados ausentes em X_1 (Rubin [14]).

Recomenda-se que os modelos de imputação incorporem ainda características de desenhos amostrais complexos como, por exemplo, pesos e estratos amostrais, por meio da inclusão de variáveis indicadoras dos estratos como covariáveis e conglomerados, pela utilização de um modelo de imputação com efeitos aleatórios no nível do conglomerado [13,15].

Quando as perdas ocorrem em múltiplas variáveis, o processo de imputação torna-se mais complexo, já que as covariáveis utilizadas no modelo de imputação podem também apresentar perdas. Um dos métodos utilizados nesse caso é a imputação pela especificação condicional completa, que realiza as imputações por um processo iterativo de regressões. Uma discussão completa do método é apresentada em van Buuren et al. [16].

- **Análise de Casos Completos: regressão logística com covariável incompleta**

A validade do ajuste do modelo de regressão logística com dados ausentes nas covariáveis pelo método de ACC depende de suposições distintas daquelas do método de IM. Considere novamente um modelo logístico no qual a perda ocorre na covariável X_1 e que as demais covariáveis $X_{2:k}$ e a variável dependente *Y* são completamente observadas. Em termos práticos, a restrição das análises a indivíduos com dados completos (ACC) seleciona uma subamostra dos participantes do estudo. No mecanismo de PCA, essa subamostra é representativa da população e assim as estimativas da ACC são válidas.

A ACC pode selecionar uma subamostra com características distintas da amostra total do estudo quando as perdas em X_1 dependem dos valores de *Y*. Por exemplo, indivíduos com dados completos podem apresentar estimativas da OR distintas daqueles omitidos das análises e, portanto, distintas das estimativas que seriam obtidas caso não houvesse perdas. Isso pode ser visto como uma interação entre as perdas e o efeito das covariáveis. A ACC pode ainda, como apontado por Vach & Illi [17], criar interações entre as próprias covariáveis do modelo logístico, anteriormente inexistentes. Como consequência, um modelo que seria bem ajustado sem essas interações caso as perdas não ocorressem pode deixar de sê-lo na ACC.

Ainda com relação aos aspectos teóricos, a ACC deve apresentar resultados válidos se a probabilidade de perda em X_1 para indivíduos com o mesmo perfil nas covariáveis (X_1 e $X_{2:k}$) não está associada a *Y*. Isso ocorre mesmo que a probabilidade de perda dependa dos próprios valores de X_1 (mecanismo de PNA) e de $X_{2:k}$. Existem ainda condições específicas de dependência entre a probabilidade de perda e *Y*, em que ACC resulta em viés apenas para o intercepto do modelo [17].

O inquérito *Saúde em Beagá*

Dados deste estudo são provenientes do projeto SBH, um inquérito de saúde domiciliar realizado pelo Observatório de Saúde Urbana de Belo Horizonte (OSUBH), da Universidade Federal de Minas Gerais (UFMG), em dois dos nove distritos sanitários de Belo Horizonte no biênio 2008-2009. Foi adotada uma amostra estratificada por conglomerados em três estágios. O índice de vulnerabilidade à saúde [18], um indicador socioeconômico geocodificado por setor censitário, foi utilizado para subdividir a área da pesquisa em três estratos amostrais. Dentro de cada estrato foram selecionados: (a) setor censitário (um total de 150 nos três estratos), com probabilidades de seleção distintas que visaram a atender objetivos secundários do projeto, ampliando a amostra em áreas de interesse; (b) domicílio, selecionado por meio de amostra aleatória simples dos domicílios cadastrados na base de dados da Prefeitura Municipal de Belo Horizonte; (c) 1 morador adulto (18 anos ou mais) e 1 morador na faixa de 11 a

17 anos, ambos selecionados aleatoriamente no domicílio.

Os adultos, foco deste estudo, após a entrevista face a face tiveram aferidas suas medidas de peso (balança Tanita BC-553, Tanita Corporation of America Inc., Arlington Heights, Estados Unidos) e altura (estadiômetro móvel), que permitiram o cálculo do IMC_direto. O IMC_referido foi calculado pelas medidas de peso e altura obtidas respectivamente pelas perguntas: "*O(A) Sr.(a) sabe seu peso (mesmo que seja valor aproximado)?*" e "*O(A) Sr.(a) sabe sua altura?*". Da amostra final de 4.048 adultos, foram excluídas previamente das análises as mulheres grávidas (n = 47) para evitar distorções nas associações do IMC com outras variáveis, e os indivíduos com o IMC_direto ausente (n = 11), resultando numa base final de 3.990 adultos (1.653 homens e 2.337 mulheres), todos com medidas do IMC_direto.

Metodologia de validação das análises

A variável IMC_referido estava ausente para 789 indivíduos (21%), o que ocorreu principalmente devido a não resposta da medida de altura (15%). Optou-se por utilizar nas análises deste artigo apenas a variável IMC_direto, com perdas que foram geradas artificialmente para refletir as perdas do IMC_referido. Assim, os valores do IMC_direto foram excluídos para os mesmos 789 indivíduos com dados referidos ausentes, criando uma nova variável pós-exclusão, que será diferenciada por um asterisco (IMC_direto*). Em seguida, os dados do IMC_direto* foram analisados pelos métodos de ACC e IM, e os resultados foram comparados com aqueles obtidos com o IMC_direto pré-exclusão. Essa abordagem preserva o mecanismo de perda original da variável IMC_referido e permite uma validação consistente.

Variáveis e modelos logísticos avaliados

Para avaliar possíveis desvios nas estimativas dos coeficientes de regressão logística quando o IMC_direto* é uma das covariáveis, foram testados três modelos que se diferenciam pela variável dependente: (1) Diabetes (1.sim, 0.não); (2) Hipertensão (1.sim, 0.não); (3) Peso_acima (1.acima do peso, 0.satisfeito ou abaixo). As duas primeiras variáveis dependentes foram obtidas de uma questão sobre doenças crônicas ("*Alguma vez, um médico ou outro profissional de saúde já disse que o(a) Sr.(a) tem alguma dessas doenças crônicas listadas abaixo?*") e a última, da pergunta: "*Com relação a seu peso, o(a) Sr.(a) está:*", que foi tratada de forma dicotômica. Além do IMC_direto*, as covariáveis Idade, Cor e Escolaridade, categorizadas em três níveis (Tabela 2), compõem os modelos logísticos. As análises foram estratificadas por sexo, resultando, assim, em 6 modelos logísticos.

As estimativas de cada modelo logístico com a covariável IMC_direto servirão de referência para a comparação com as estimativas obtidas com a mesma variável pós-exclusão. O modelo de referência pode ser visto como uma ACC com perdas mínimas, já que no máximo 18 mulheres e 8 homens são ignorados por não apresentarem dados completos nas variáveis de cada modelo.

Análise dos dados

Para o ajuste dos modelos logísticos com a covariável IMC_direto*, além da ACC, foram avaliadas quatro abordagens de IM (IM0, IM1, IM2 e IM3), que se diferenciam unicamente pelas variáveis utilizadas no modelo de imputação. A ACC reduz sensivelmente a amostra, ignorando os 789 indivíduos (586 mulheres e 203 homens) com valores ausentes no IMC_direto*. Nas abordagens de IM, as variáveis incluídas no modelo de imputação foram selecionadas de forma cumulativa, assim, o modelo de IM1 contém todas as variáveis utilizadas no IM0 e assim sucessivamente.

O modelo de imputação de IM0 inclui apenas as covariáveis dos modelos logísticos (Idade, Cor e Escolaridade) e, assim, omite a variável dependente. Espera-se aqui ilustrar o potencial viés dessa abordagem. O IM1 contém as mesmas variáveis utilizadas em cada modelo de análise, acrescentado comparativamente a IM0, a variável dependente de cada modelo logístico. As variáveis adicionadas aos modelos IM2 e IM3 foram selecionadas seguindo a estratégia de van Buuren et al. [12] descrita anteriormente.

O modelo IM2 inclui, além das variáveis do modelo de análise, as 8 principais variáveis associadas ao IMC_direto* ou à variável indicadora das perdas do IMC_direto* (1.ausente, 0.observado) e pode ser visto, segundo discutido por Collins et al. [8], como uma abordagem restritiva de seleção de variáveis. O modelo IM3 adiciona ainda 8 variáveis e pode ser classificado como uma abordagem inclusiva, já que as variáveis incluídas têm menos poder de predição ou associação mais fraca com as perdas. Todas as variáveis incluídas em IM2 e IM3 tiveram associação estatisticamente significativa com o IMC_direto* ou com a variável indicadora de perdas.

Dentre as 8 variáveis incluídas no IM2 destacam-se as que atuam como *proxy* do IMC: Silhueta, Silhueta2 e Satisfação_corporal. A variável Silhueta tem valores de 1 a 9 e corresponde à escala de silhuetas de Stunkard et al. [19], discutida

Tabela 2

Distribuição das variáveis utilizadas nos modelos de regressão, proporção de indivíduos com IMC_referido ausente e teste para diferença nessas proporções segundo sexo.

Variáveis	Feminino			Masculino		
	n	IMC_referido ausente (%)	Valor de p	n	IMC_referido ausente (%)	Valor de p
IMC_direto						
< 25	1.043	22,8	0,01	817	13,7	0,16
≥ 25 e < 30	759	23,7		592	10,3	
≥ 30	535	31,4		244	12,3	
Cor						
Branca	933	21,3	< 0,01	593	11,1	0,39
Negra	274	36,5		241	14,5	
Outra	1.117	25,2		814	12,4	
Escolaridade (anos)						
0-3	395	49,9	< 0,01	195	22,6	< 0,01
4-10	913	29,0		692	14,7	
11+	1.027	12,1		766	7,4	
Idade (anos)						
18-39	960	22,9	< 0,01	750	14,1	0,01
40-59	875	23,5		578	9,0	
60+	502	31,9		325	13,8	
Variáveis dependentes dos modelos logísticos						
Peso_acima						
Satisfeito ou abaixo	1.021	29,3	< 0,01	1.088	13,7	0,01
Acima	1.313	21,7		565	9,6	
Diabetes						
Não	2.133	24,6	0,05	1.533	12,3	0,93
Sim	202	30,7		117	12,0	
Hipertensão						
Não	1.566	22,3	< 0,01	1.186	12,2	0,90
Sim	769	30,7		466	12,4	

IMC: índice de massa corporal.

e ilustrada mais recentemente por Gardner et al. [20]. A escala é composta por 9 figuras masculinas e femininas variando nos extremos, da magreza à representação de um indivíduo obeso. A variável Silhueta corresponde à escolha do entrevistado sobre a "figura que se parece mais com" ele hoje, e a variável Silhueta2 corresponde a essa medida elevada ao quadrado. A variável Satisfação_corporal representa a diferença entre a Silhueta e a escolha da figura que o indivíduo gostaria de se parecer.

As demais variáveis do modelo IM2 são listadas a seguir com uma breve descrição entre parênteses: Estado_civil (4 categorias: solteiro; casado/amigado; separado/desquitado; viúvo), Renda_familiar (4 categorias em salários mínimos: < 2 SM; 2 < 5 SM; 5 < 10 SM; 10+ SM), Doença_mental (proveniente da questão sobre doenças crônicas), Mudança_peso (está tentando alterar o peso), e Inativo (não praticou atividade física nos últimos 3 meses). As 8 variáveis adicionadas ao IM3 são: Colesterol_alto, Artrite e Epilepsia, também provenientes da questão sobre doenças crônicas; Chefe_família (é chefe da família), Posse_veículo (possui carro ou moto), Restrição_alimentar (alguma vez na vida reduziu ou deixou de fazer refeições devido a problemas financeiros), Consumo_fruta (consumo de frutas semanal, 3 categorias) e Fumo (fuma diariamente).

Embora no modelo logístico o IMC seja tratado como variável categórica, a imputação foi realizada considerando sua forma contínua por meio de um modelo linear. Após avaliações, veri-

ficou-se que a transformação logarítmica da variável IMC_direto* foi a que apresentou melhor ajuste ao modelo linear e, portanto, a transformação foi aplicada previamente às imputações. Após as imputações, aplicou-se a transformação inversa e a variável foi categorizada. Todas as imputações levaram em consideração os pesos amostrais da pesquisa e incluíram ainda uma variável identificadora dos estratos amostrais. Em todas as abordagens foram realizadas 20 imputações independentemente para cada sexo, acompanhando a estratificação das análises.

O teste Kolmogorov-Smirnov foi utilizado para avaliar a igualdade da distribuição da variável IMC_direto entre os grupos de indivíduos com IMC_referido observado e ausente, como abordagem para checar a suposição de PNA, e o teste qui-quadrado, para verificar associações entre as perdas e variáveis utilizadas nos modelos logísticos. Com o objetivo de se verificar problemas resultantes da seleção de indivíduos da ACC, o teste Wald para múltiplos coeficientes foi utilizado para testar a significância conjunta das interações entre as perdas do IMC_referido e os coeficientes de cada modelo logístico de referência. Avaliou-se ainda a qualidade do ajuste dos modelos logísticos de referência (pré-exclusão) e dos modelos da ACC (pós-exclusão) por meio do teste de Pearson. Foi adotado o nível de significância de 5% em todos os testes.

Todas as análises foram realizadas no *software* Stata 11 (Stata Corp., College Station, Estados Unidos) e o comando "svy" permitiu considerar o desenho amostral nos ajustes dos modelos logísticos segundo as abordagens de ACC (comando "svy:logit") e IM (ver Royston et al. [21] para o comando "mim:svy:logit", utilizado para o ajuste nos bancos pós-imputação). O comando "svy:" permite corrigir as estimativas dos coeficientes e respectivos erros padrão dos modelos logísticos, ao considerar as complexidades do desenho amostral (estratos e conglomerados) e ainda as diferentes probabilidades de seleção dos indivíduos, ocasionadas pelo desenho de múltiplos estágios e sobreamostragem de determinadas regiões. As análises relativas à distribuição das perdas não estão ponderadas, e assim as proporções de perdas apresentadas se referem ao que efetivamente foi observado no estudo.

As imputações deste artigo foram realizadas utilizando-se o comando "mi ice" (ver Royston [22] para detalhes da última atualização da implementação) que utiliza como referência o método descrito por van Buuren et al. [12] (ver van Burren & Groothuis-Oudshoorn [23] para uma implementação alternativa da IM no *software* gratuito R, no pacote MICE). O comando "mi ice" tem como limitação a impossibilidade de considerar no ajuste dos modelos de imputação o efeito de conglomerados amostrais.

Resultados

O nível de perdas na variável IMC_referido foi quase duas vezes superior no sexo feminino (25,1% contra 12,3% do sexo masculino), e assim espera-se observar para este grupo um maior impacto nas análises. A distribuição do IMC_direto das mulheres com IMC_referido ausente apresentou caudas mais largas do que daquelas com IMC_referido observado ($p < 0,01$; Figura 1). Assim, mulheres com valores extremos do IMC_direto têm maior propensão a perdas, isto é, não responder às medidas referidas de peso e/ou altura. Homens com IMC_referido ausente tiveram menores valores do IMC_direto, dado o deslocamento da distribuição para a esquerda ($p < 0,01$). Isso indica que controlando apenas pela variável Sexo, o mecanismo de PNA não pode ser descartado.

Os dados da Tabela 2 apresentam maiores detalhes da distribuição das perdas na variável IMC_referido. Indivíduos mais escolarizados apresentaram menor proporção de perdas com forte gradação para ambos os sexos. A perda, diferentemente do que foi verificado para a variável contínua, está associada ao IMC_direto na sua forma categorizada apenas para o sexo feminino. Entre as mulheres, a perda está associada a todas as variáveis dependentes dos modelos logísticos. Para os homens, a proporção de perdas é praticamente constante nos níveis das variáveis dependentes Diabetes e Hipertensão, mas está associada à variável Peso_acima.

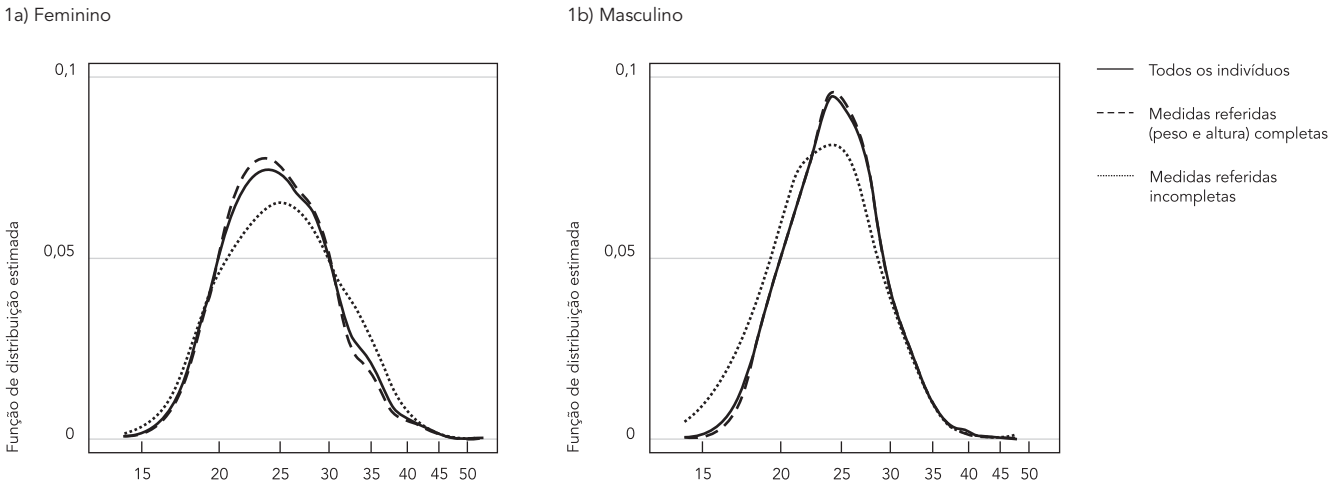
Os resultados dos ajustes dos modelos logísticos utilizando-se a variável IMC_direto (pré-exclusão), que servem como referência para a avaliação dos métodos de ACC e IM, são apresentados na Tabela 3. Considerando os 6 modelos ajustados, temos um total de 48 coeficientes estimados (sem levar em conta os 6 interceptos), dos quais 32 são estatisticamente significativos.

Testes indicaram bom ajuste de todos os modelos logísticos de referência (Tabela 3) com exceção do modelo Hipertensão para o sexo feminino ($p = 0,01$). A inclusão de interações da covariável IMC_direto com as demais melhoraram o ajuste deste modelo ($p = 0,37$; resultados não mostrados, ver Discussão), o mesmo foi verificado na ACC. O modelo Diabetes para o sexo feminino, que na análise de referência apresentou bom ajuste ($p = 0,11$), foi o único que na ACC perdeu esta qualidade ($p = 0,03$).

Para cada modelo de referência (Tabela 3) foi avaliada a possibilidade de interação entre os

Figura 1

Distribuição do IMC_direto para todos os indivíduos, para aqueles com medidas referidas de peso e altura completas e para aqueles com medidas incompletas, segundo sexo.



IMC: índice de massa corporal.
Nota: valores do IMC em escala logarítmica.

Tabela 3

Resultados dos modelos logísticos com a covariável IMC_direto (pré-exclusão) segundo variável dependente e sexo.

Covariável	Diabetes				Hipertensão				Peso_acima			
	Feminino (n = 2.320)		Masculino (n = 1.645)		Feminino (n = 2.320)		Masculino (n = 1.647)		Feminino (n = 2.319)		Masculino (n = 1.648)	
	OR	IC95%	OR	IC95%	OR	IC95%	OR	IC95%	OR	IC95%	OR	IC95%
IMC_direto												
< 25 (ref.)	1,0		1,0		1,0		1,0		1,0		1,0	
≥ 25 e < 30	**1,7**	1,0-2,8	1,8	1,0-3,5	**2,0**	1,5-2,8	**1,7**	1,1-2,4	**13,6**	9,9-18,6	**16,7**	11,2-25,0
≥ 30	**2,8**	1,7-4,8	**2,8**	1,3-5,7	**4,9**	3,5-7,0	**3,8**	2,4-6,2	**40,3**	24,8-65,3	**88,1**	48,9-158,5
Cor												
Branca (ref.)	1,0		1,0		1,0		1,0		1,0		1,0	
Negra	1,6	0,9-3,0	0,8	0,3-2,2	**2,6**	1,7-3,9	**2,4**	1,4-4,1	0,7	0,5-1,1	**0,5**	0,3-0,8
Outra	**1,8**	1,2-2,7	1,2	0,7-2,0	1,3	1,0-1,7	1,1	0,7-1,6	0,9	0,6-1,2	**0,6**	0,4-0,9
Escolaridade (anos)												
0-3 (ref.)	1,0		1,0		1,0		1,0		1,0		1,0	
4-10	1,0	0,7-1,6	1,3	0,5-3,2	**0,6**	0,4-0,9	**0,5**	0,3-0,7	**1,8**	1,2-2,7	1,3	0,7-2,6
11+	0,6	0,3-1,0	1,5	0,6-3,3	**0,4**	0,3-0,6	**0,4**	0,3-0,7	**3,6**	2,4-5,4	**2,6**	1,3-5,2
Idade (anos)												
18-39 (ref.)	1,0		1,0		1,0		1,0		1,0		1,0	
40-59	**3,7**	2,0-6,9	**8,7**	2,9-26,3	**4,8**	3,4-6,9	**2,7**	1,7-4,2	0,7	0,5-1,0	1,0	0,7-1,4
60+	**12,0**	6,7-21,7	**26,6**	9,1-77,6	**22,5**	15,4-32,8	**9,8**	5,9-16,4	**0,2**	0,1-0,3	0,6	0,4-1,1

IC95%: intervalo de 95% de confiança; IMC: índice de massa corporal; OR: *odds ratio*; ref.: referência.
Nota: valores da OR em negrito indicam p < 0,05.

efeitos das covariáveis e uma variável indicadora das perdas no IMC_referido. Apenas no modelo Peso_acima (sexos masculino e feminino), a inclusão de interações para todas as covariáveis (além da própria variável indicadora) foi estatisticamente significativa, indicando que os indivíduos com dados completos apresentam associações (entre covariáveis e variável dependente) distintas dos demais (IMC_referido ausente).

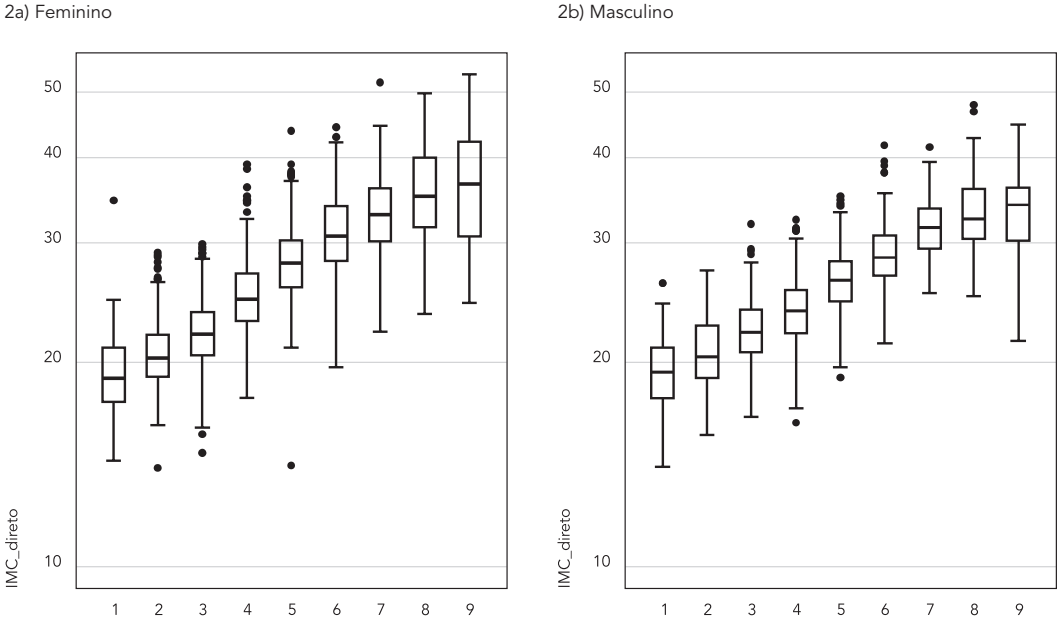
No processo de seleção de variáveis para os modelos de imputação de IM2 e IM3, a variável Silhueta foi a mais importante preditora do IMC. O ajuste de um modelo linear do logaritmo do IMC_direto* tendo como covariáveis apenas a variável Silhueta e a interação da mesma com a variável Sexo, apresentou 60% da variância explicada pelo modelo (R^2). Esse percentual aumentou para 62% quando a variável Satisfação_corporal (categorizada) e a interação da mesma com a variável Sexo foram incluídas no modelo. A Figura 2 ilustra a relação quase linear entre a variável Silhueta e a mediana do IMC_direto (na escala logarítmica).

Após a exclusão dos valores do IMC_direto e a aplicação dos métodos de ACC e de IM, os resultados de cada modelo logístico foram comparados ao modelo de referência. Como esperado, na abordagem IM0, que omite a variável dependente do modelo de imputação, a OR da variável IMC_direto* foi subestimada em todos os modelos, reduzindo para menos da metade da OR de referência no modelo Peso_acima para o sexo feminino (Figura 3). De modo geral, o desvio de cada estimativa da OR reduz de forma consistente quando se caminha, na Figura 3, da ACC para a abordagem IM2 e se mantém constante na abordagem IM3.

Similarmente à Figura 3, a Figura 4 apresenta a média e amplitude da razão entre a OR estimada em cada abordagem e a OR de referência. Entretanto, nessa, as razões menores do que 1 são invertidas e os desvios da OR podem ser avaliados independentemente de se tratarem de sub ou sobre-estimações. Também são omitidos todos os valores referentes às estimativas do intercepto do modelo logístico, que geralmente

Figura 2

Box-plot da variável IMC_direto segundo a variável Silhueta e Sexo.



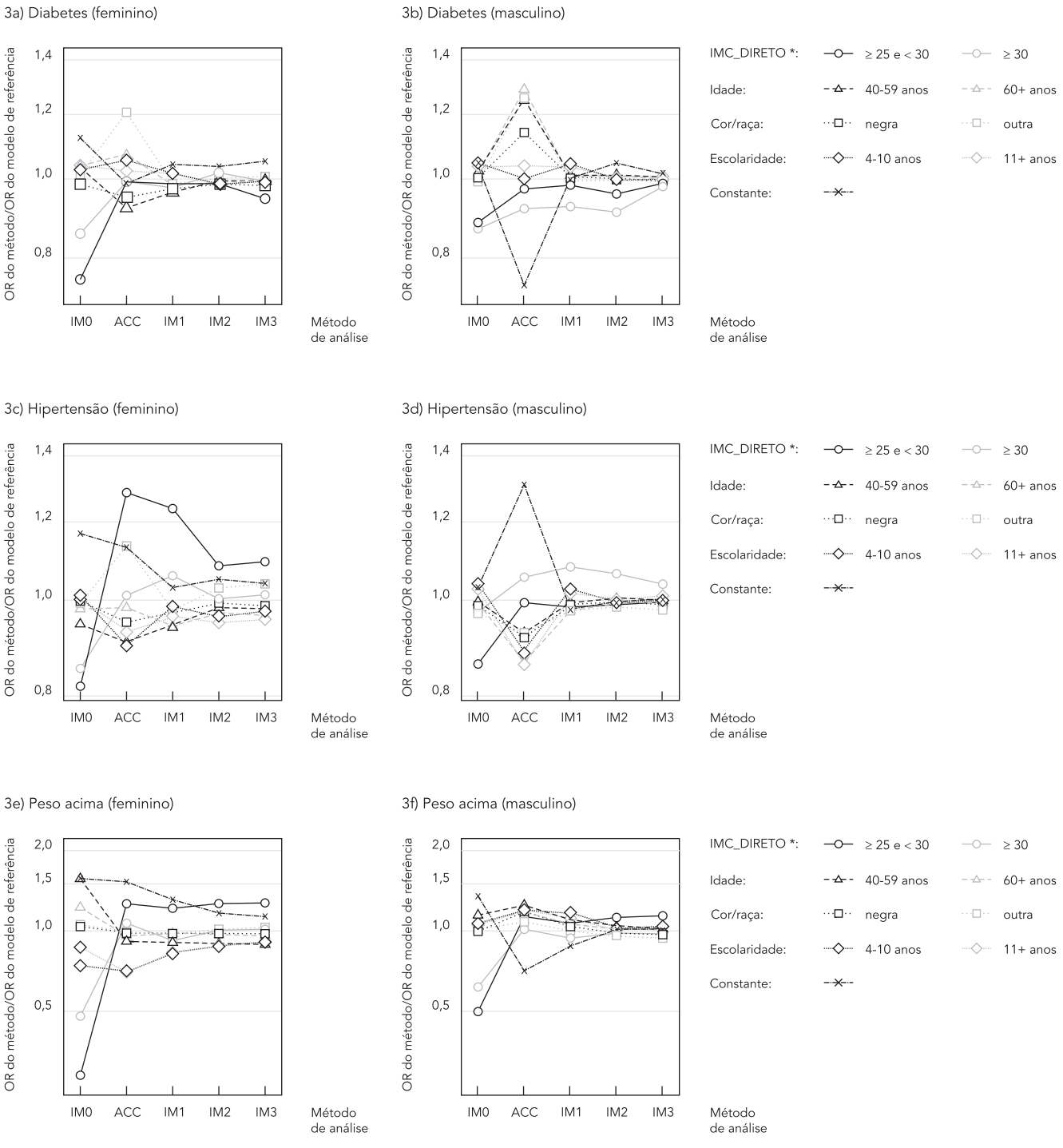
IMC: índice de massa corporal.

Nota: valores do IMC em escala logarítmica.

Silhueta: Figura com que mais se parece hoje (1 = magro, 9 = obeso).

Figura 3

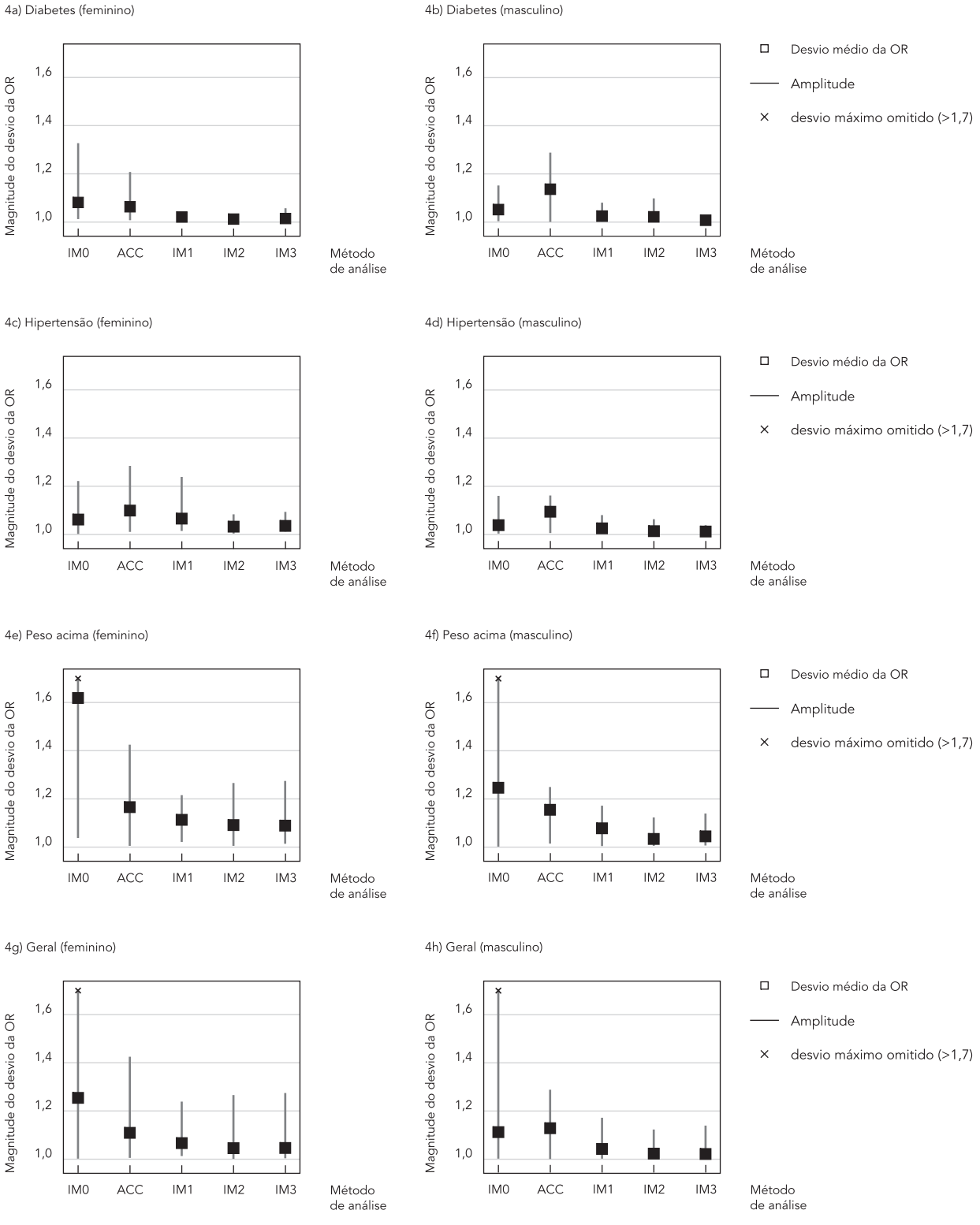
Razão (em escala logarítmica) entre a OR estimada nas abordagens de ACC e IM e a OR do modelo de referência (pré-exclusão) segundo a variável dependente e sexo.



ACC: Análise de Casos Completos; IM: Imputação Múltipla; IMC: índice de massa corporal; OR: *odds ratio*.

Figura 4

Média e amplitude do desvio entre a OR estimada nas abordagens de ACC e IM e a OR do modelo de referência (pré-exclusão) segundo a variável dependente e sexo.



ACC: Análise de Casos Completos; IM: Imputação Múltipla; OR: *odds ratio*.

não são de interesse do pesquisador. As Figuras 4g e 4h apresentam um resumo dos resultados por sexo.

Nos resultados gerais de ambos os sexos, a ACC apresentou em média o maior desvio na OR dentre as abordagens consideradas próprias (feminino: ACC com desvio médio de 11%, IM1 6,7%, IM2 4,6% e IM3 4,7%; masculino: ACC 12,9%, IM1 4,3%, IM2 2,3%, IM3 2,2%). Esse pior desempenho da ACC também é verificado em cada um dos modelos logísticos avaliados. Observa-se em geral um melhor desempenho das abordagens de IM entre os homens, nos quais houve uma menor proporção de perdas do IMC_referido. O modelo IM0, sem surpresas, foi a abordagem que apresentou o maior desvio médio da OR entre as mulheres e também a maior variação (amplitude) para ambos os sexos.

É natural que o erro padrão de um coeficiente estimado pelos métodos de análise de dados incompletos seja superior àquele que idealmente seria obtido se todos os dados fossem observados. Entretanto, métodos mais eficientes tendem a produzir erros padrão próximos desse ideal. Os erros padrão dos coeficientes estimados nas abordagens de_ACC, IM1, IM2 e IM3 apresentaram respectivamente inflação média de 18,9%, 5,6%, 3,2% e 3,5%, para o sexo feminino, e de 7%, 2,7%, 1,4% e 1,9% para o masculino quando comparados aos erros padrão respectivos do modelo de referência (Figura 5 – resultados gerais). O elevado percentual de perdas do IMC_referido no sexo feminino teve impacto sensível na inflação dos erros padrão deste grupo.

A abordagem IM0, que claramente vicia a OR da variável IMC_direto*, não incorpora essa imprecisão no erro padrão de seus coeficientes. Isso indica que os intervalos de confiança produzidos por essa abordagem são inconsistentes, isto é, não devem refletir a confiança nominal (95%).

A inflação do erro padrão reflete diretamente na significância dos coeficientes. Dos 32 coeficientes significantes dos modelos de referência (Tabela 3), deixaram de ser significativos a OR do IMC_direto (≥ 25 e < 30) do modelo Diabetes (feminino) em todos os métodos, a OR da Escolaridade (4-10 anos) do modelo Peso_acima (feminino) na ACC e IM1, e a OR da Cor (outra) do modelo Peso_acima (masculino) apenas na ACC. Dos 16 coeficientes não significativos, apenas a abordagem IM3 apresentou divergência (OR da categoria "40-59 anos" do modelo Peso_acima, sexo feminino). Assim, a ACC foi o método com maior discrepância em relação aos modelos de referência (3 resultados divergentes).

Discussão

Em todos os modelos logísticos avaliados, as abordagens de imputação IM1, IM2 e IM3 tiveram, em média, menor desvio da OR em relação às estimativas de referência (pré-exclusão) quando comparadas com o desvio da ACC. Assim, mesmo a abordagem de imputação mais simples (IM1), que inclui no modelo de imputação apenas as variáveis do modelo de análise, obteve melhor desempenho que a ACC. Além disso, os intervalos de confiança da ACC apresentaram uma inflação consideravelmente maior que todas as abordagens de IM, levando a não detecção de efeitos que, no modelo de referência, eram significativos. As abordagens IM2 e IM3 que utilizam a estratégia de seleção de variáveis proposta por van Buuren et al. [12] tiveram menor desvio da OR e inflação do erro padrão que a abordagem IM1.

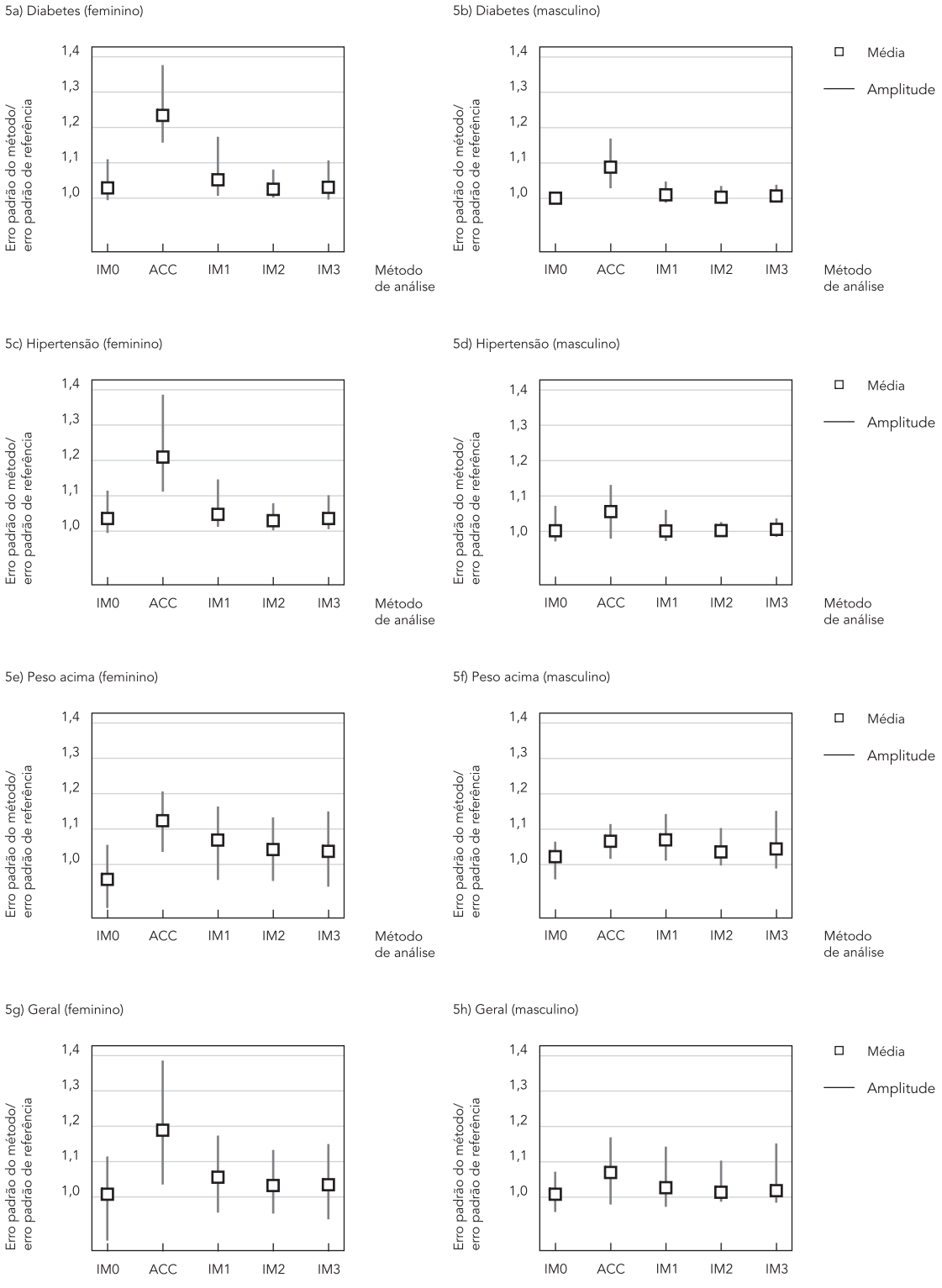
Os piores resultados da ACC indicam a existência de algum nível de associação entre as perdas no IMC_referido e as variáveis dependentes avaliadas, inclusive para indivíduos com o mesmo perfil nas covariáveis [7,17]. Isso fez com que as pessoas com IMC_referido observado (dados completos) tivessem relações entre covariáveis e variável dependente distintas das demais (modelo Peso_acima, sexos masculino e feminino). No modelo Diabetes para o sexo feminino, observou-se ainda que a restrição das análises aos indivíduos com dados completos levou a um modelo mal ajustado, uma possibilidade apontada por Vach & Illi [17].

Neste estudo, a abordagem inclusiva de seleção de variáveis para o modelo de imputação, representada por IM3, teve resultados muito similares ao obtido pela abordagem restritiva (IM2). Segundo Collins et al. [8], que avaliam por meio de simulações diferentes abordagens de imputação em uma única variável (X_1), estratégias inclusivas são preferíveis, pois reduzem o risco de omissão de variáveis importantes no modelo de imputação e, portanto, o risco de viés. Verificou-se, por exemplo, que a não inclusão de uma variável sabidamente associada a essas perdas no modelo de imputação levou a viés de estimativas posteriores, e ainda que a magnitude do viés depende do grau de correlação dessa variável omitida com X_1. Em outra circunstância, na qual foi imposta uma PNA em X_1, verificou-se que a inclusão de variáveis fortemente relacionadas a X_1 no modelo de imputação resultou em redução considerável no viés das análises posteriores.

Nunes et al. [24] avaliaram o impacto de perdas em uma única covariável simulada sob o mecanismo de PCA, e também verificaram que os intervalos de confiança da abordagem de IM são mais próximos dos obtidos com o banco comple-

Figura 5

Média e amplitude da razão entre o erro padrão dos coeficientes estimados nas abordagens de ACC e IM e o erro padrão estimado pelo modelo de referência (pré-exclusão) segundo a variável dependente e sexo.



ACC: Análise de Casos Completos; IM: Imputação Múltipla.

to do que aqueles da ACC. Mais recentemente, Nunes et al. [25] discutem e comparam métodos de imputação simples e múltipla numa reanálise dos dados do estudo anterior, agora preservando as perdas originais da covariável. Distintamente de ambos, o presente estudo é mais abrangente, já que alia a preservação do mecanismo de perda original da variável IMC_referido à possibilidade de validação dos resultados.

Os modelos apresentados neste estudo têm como único objetivo ilustrar e comparar os métodos de análises de dados incompletos. Portanto, não houve qualquer intenção de ajustar modelos baseados em marcos teóricos, ou, por exemplo, preocupações relativas a efeitos de confusão. Considerou-se um mínimo de plausibilidade na composição dos modelos para que parte dos coeficientes tivesse significância estatística, e ainda foi avaliada a qualidade do ajuste dos mesmos. O único modelo mal-ajustado, Hipertensão para o sexo feminino, produziu resultados muito similares àquele que incluía as interações necessárias. Assim, para evitar uma apresentação demasiadamente complexa nos gráficos e tabelas, optou-se pela publicação dos resultados sem interações.

É importante ressaltar que embora as perdas no IMC_direto tenham sido geradas artificialmente, reproduzindo as perdas do IMC_referido, este estudo busca aproximar ao máximo a situação em que os dados ausentes do IMC_referido são recuperados. Entendemos que os resultados observados aqui são generalizáveis para estudos que trabalham unicamente com a variável IMC_referido, e recomendamos que o método de IM seja preferido em relação a ACC nas análises que envolvam esta variável.

A recuperação de dados ausentes pode trazer diversas dificuldades práticas, já que os problemas que levaram à ausência de informações podem se repetir (e.g. recusas), ou ainda esta recuperação pode ser inviável, por exemplo, no caso de perda de seguimento em um estudo de coorte. A abordagem aqui adotada permite, de forma simples e eficaz, avaliar o impacto real das perdas nas análises. Sugerimos a condução de outros estudos similares para a acumulação de novas evidências.

Resumo

Pesquisadores da área da saúde lidam frequentemente com o problema das bases de dados incompletas. A Análise de Casos Completos (ACC), que restringe as análises aos indivíduos com dados completos, reduz o tamanho da amostra e pode produzir estimativas viciadas. Baseado em fundamentos estatísticos, o método de Imputação Múltipla (IM) utiliza todos os dados coletados e é recomendado como alternativa à ACC. Dados do estudo Saúde em Beagá*, inquérito domiciliar em que participaram 4.048 adultos de dois dos nove distritos sanitários da Cidade de Belo Horizonte no biênio 2008-2009, foram utilizados para avaliar a ACC e diferentes abordagens de IM no contexto de modelos logísticos com covariáveis incompletas. Peculiaridades de algumas variáveis desse estudo permitiram aproximar uma situação em que os dados ausentes de uma covariável são recuperados, e assim os resultados anteriores e posteriores à recuperação são comparados. Verificou-se que mesmo a abordagem mais simplista de IM obteve melhor desempenho que a ACC, já que se aproximou mais dos resultados pós-recuperação.*

Métodos; Análise Estatística; Modelos Logísticos; Índice de Massa Corporal

Colaboradores

V. P. Camargos, C. C. César e F. A. Proietti contribuíram com o delineamento do estudo, a análise dos dados, redação do trabalho, revisão crítica e aprovação do manuscrito final. W. T. Caiaffa e C. C. Xavier contribuíram com o delineamento do estudo, a coleta de dados, revisão crítica e aprovação do manuscrito final.

Agradecimentos

Agradecemos a todos os membros do Observatório de Saúde Urbana de Belo Horizonte que participaram do projeto *Saúde em Beagá,* e o apoio da Secretaria Municipal de Saúde de Belo Horizonte na coleta de dados. O projeto foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundo Nacional de Saúde do Ministério da Saúde e Fundação de Amparo à Pesquisa do Estado de Minas Gerais. C. C. César, F. A. Proietti e W. T. Caiaffa recebem bolsa de produtividade em pesquisa do CNPq.

Referências

1. Raghunathan TE. What do we do with missing data? Some options for analysis of incomplete data. Annu Rev Public Health 2004; 25:99-117.
2. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002; 7: 147-77.
3. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. Am Stat 2007; 61:79-90.
4. World Health Organization. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser 2000; 894:i-xii, 1-253.
5. Rubin DB. Inference and missing data. Biometrika 1976; 63:581-90.
6. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009; 338:b2393.
7. Little RJ. Regression with missing X's: a review. J Am Stat Assoc 1992; 87:1227-37.
8. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and 24 restrictive strategies in modern missing data procedures. Psychol Methods 2001; 6:330-51.
9. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol 2009; 60: 549-76.
10. Glynn RJ, Laird NM, Rubin DB. Multiple imputation in mixture models for nonignorable nonresponse with follow-ups. J Am Stat Assoc 1993; 88:984-93.
11. Rubin DB. Multiple imputations in sample surveys – a phenomenological bayesian approach to nonresponse. In: Proceedings of the Survey Research Methods Section, Am Stat Assoc. http://www.amstat.org/sections/srms/proceedings/papers/1978_004.pdf (acessado em 18/Jan/2011).
12. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999; 18:681-94.
13. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996; 91:473-89.
14. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
15. Little RJ, Rubin DB. Statistical analysis with missing data. 2nd Ed. New York: John Wiley & Sons; 2002.
16. van Buuren S, Brand JP, Groothius-Oudshoorn CG, Rubin DB. Fully conditional specification in multivariate imputation. J Stat Comput Simul 2006; 76:1049-64.
17. Vach W, Illi S. Biased estimation of adjusted odds ratios from incomplete covariate data due to violation of the missing at random assumption. Biom J 1997; 39:13-28.
18. Gerência de Epidemiologia e Informação, Secretaria Municipal de Saúde de Belo Horizonte. Índice de vulnerabilidade à saúde, 2003. http://www.pbh.gov.br/smsa/biblioteca/gabinete/risco2003 (acessado em 18/Jan/2011).
19. Stunkard AJ, Sorensen T, Schulsinger F. Use of the Danish adoption registry for the study of obesity and thinness. In: Skety S, Rowland LP, Sidman RL, Matthysse SW, editors. Genetics of neurological and psychiatric disorders. New York: Raven Press; 1983. p. 115-20.
20. Gardner RM, Friedman BN, Jackson NA. Methodological concerns when using silhouettes to measure body image. Percept Mot Skills 1998; 86: 387-95.
21. Royston P, Carlin JB, White IR. Multiple imputation of missing values: new features for mim. Stata Journal 2009; 9:252-64.
22. Royston P. Multiple imputation of missing values: further update of ice, with an emphasis on categorical variables. Stata J 2009; 9:466-77.
23. van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. http://www.stefvanbuuren.nl/publications/MICE%20in%20R%20-%20Draft.pdf (acessado em 17/Jun/2011).
24. Nunes LN, Klück MM, Fachel JMG. Uso da imputação múltipla de dados faltantes: uma simulação utilizando dados epidemiológicos. Cad Saúde Pública 2009; 25:268-78.
25. Nunes LN, Klück MM, Fachel JMG. Comparação de métodos de imputação única e múltipla usando como exemplo um modelo de risco para mortalidade cirúrgica. Rev Bras Epidemiol 2010; 13:596-606.

Recebido em 19/Jan/2011
Versão final reapresentada em 18/Jun/2011
Aprovado em 10/Ago/2011