

Betzabeth Slater

Dirce Maria Lobo Marchioni

Silvia Maria Voci

# Aplicação de regressão linear para correção de dados dietéticos

## Use of linear regression for correction of dietary data

---

### RESUMO

**OBJETIVO:** Estratégias metodológicas vêm sendo desenvolvidas para minimizar o efeito do erro de medida da dieta. O objetivo do estudo foi descrever a aplicação de uma estratégia para correção da informação dietética pelo erro de medida.

**MÉTODOS:** Foram obtidos dados de consumo alimentar pela aplicação do Questionário de Frequência Alimentar a 79 adolescentes do Município de São Paulo em 1999. Os dados dietéticos obtidos foram corrigidos por meio de regressão linear, após o ajuste pela energia usando-se o método dos resíduos. O método de referência utilizado foi o recordatório de 24 horas, aplicado em três momentos distintos.

**RESULTADOS:** Os valores corrigidos aproximaram-se dos valores de referência. O fator de correção  $\lambda$  foi de 0,89 para energia. Para os macronutrientes, os fatores foram de 0,41; 0,22 e 0,20, para carboidratos, lipídios e proteínas, respectivamente.

**CONCLUSÕES:** As médias e desvios-padrão dos valores corrigidos denotam que houve uma correção do erro de medida. Apesar disso, debate-se o desempenho desses métodos, que são notoriamente imperfeitos quando seus pressupostos teóricos não são atendidos, o que é comum nos estudos da dieta que usam instrumentos de medida baseados no relato dos indivíduos.

**DESCRITORES:** Consumo de alimentos. Adolescente. Viés (Epidemiologia). Questionários. Medidas, métodos e teorias. Questionário de Frequência Alimentar.

---

### ABSTRACT

**OBJECTIVE:** Methodological approaches have been developed to minimize effects of measurement error in dietary intake data. The objective of the study was to apply a strategy to correct intake data according to measurement error.

**METHODS:** Intake data were obtained by applying the Food Frequency Questionnaire in 79 adolescents of São Paulo city, Brazil. Correction of dietary intake data was performed by linear regression, after energy adjustment by the residual method. The reference method used was 24-hour dietary recall and it was applied three times.

Departamento de Nutrição. Faculdade de Saúde Pública. Universidade de São Paulo. São Paulo, SP, Brasil

**Correspondência | Correspondence:**  
Betzabeth Slater  
Departamento de Nutrição - 2º andar  
Av. Dr. Arnaldo, 715 Cerqueira César  
01246-904 São Paulo, SP, Brasil  
E-mail: bslater@usp.br

Recebido: 20/4/2006  
Revisado: 18/10/2006  
Aprovado: 23/11/2006

**RESULTS:** Corrected values were similar to reference values. Correction factor  $\lambda$  was 0.89 to energy. For macronutrients, factors were 0.41, carbohydrate 0.22, and lipid and protein 0.20.

**CONCLUSIONS:** Mean and standard deviation of corrected values show a correction for the measurement error. The performance of these methods, that are imperfect, is questioned when the assumptions are not proved which is common in intake studies of measures based on the reports of individuals.

**KEYWORDS:** Food intake. Adolescent. Bias (Epidemiology). Questionnaires. Measurements, methods and theories. Food Frequency Questionnaire.

## INTRODUÇÃO

A maior limitação dos estudos epidemiológicos que investigam a relação entre dieta e doença tem sido a dificuldade de mensurar a dieta habitual de indivíduos de forma acurada e precisa. Essa limitação fez com que se ampliasse o número de estudos com o objetivo de avaliar o desempenho de instrumentos de coleta de informações dietéticas, especialmente o Questionário de Frequência Alimentar (QFA).<sup>3</sup>

Embora tenha se tornado o instrumento mais frequente em estudos em epidemiologia nutricional, o QFA também apresenta limitações relacionadas: à memória e percepção,<sup>8</sup> falta de padronização do instrumento e treinamento dos entrevistadores, falhas na estrutura do instrumento, e também, devido ao acaso.<sup>12,18</sup>

Há mais de uma década, Beaton<sup>3</sup> afirmou que “o consumo dietético não pode ser estimado sem erro e provavelmente nunca será”. Considerando-se essa afirmativa, impõe-se o entendimento do erro como um conceito estatístico, e não no sentido de equívoco na coleta de dados. De fato, é reconhecido que a mensuração da dieta habitual dos indivíduos está sujeita a erros aleatórios e sistemáticos. O erro sistemático ou viés ocorre em média para todos os indivíduos medidos. O erro aleatório decorre, principalmente, de flutuações dadas pelo dia-a-dia e varia entre indivíduos, com média igual a zero, fornecendo medidas menos precisas.<sup>2</sup>

As correlações entre as estimativas da dieta pelo QFA e por métodos de referência frequentemente se situam entre 0,3 e 0,7, o que sugere substancial erro.<sup>7</sup> A literatura científica nessa área mostra que as medidas de associação observadas nas exposições dietéticas em estudos epidemiológicos são relativamente baixas, em geral menores que 2,0. Dessa forma, pode-se deixar de observar associações devido à falta de acurácia e precisão da medida da dieta.<sup>18</sup>

Assim, estratégias metodológicas vêm sendo desenvolvidas na tentativa de estimar os parâmetros corretamente, como no caso de estudos de validação e, mais recentemente, nos estudos de calibração.

Calibração pode ser definida como a determinação da relação entre duas escalas de medidas. Essa metodologia estatística objetiva aproximar as medidas obtidas pelo QFA dos valores reais de ingestão, estimados por métodos de referência supostamente livres de viés (erros sistemáticos), não necessariamente aplicados em mais de um momento.<sup>13</sup> Dessa forma, obtêm-se valores corrigidos, parcialmente livres do erro contido no QFA.

O principal objetivo dos estudos de calibração é o uso da informação obtida para o ajuste da medida de associação que será estimada no estudo epidemiológico principal, pela correção do erro associado com o uso do QFA.<sup>4,14,18</sup> Essa abordagem é realizada a priori, a partir de modelo de regressão paramétrico ou não-paramétrico.<sup>10,13,14</sup>

Dentre os métodos paramétricos, o da calibração por modelo de regressão linear<sup>14</sup> é considerado padrão:<sup>17</sup> a estimativa da ingestão obtida pelo método de referência é modelada como uma função da estimativa de ingestão obtida pelo QFA. O modelo de regressão linear obtido no estudo de calibração pode ser usado como modelo preditor para imputação da estimativa da ingestão verdadeira, dado o valor do QFA. Nessa abordagem são incorporados tanto o erro sistemático quanto o erro aleatório na medida da dieta obtida pelo QFA.<sup>14</sup> Quando utilizados posteriormente na análise da relação entre doença e ingestão dietética no estudo principal, os valores calibrados podem eliminar ou reduzir substancialmente, erros que poderiam afetar as estimativas de risco.<sup>10,17</sup>

Assim, o objetivo do presente trabalho foi descrever a aplicação de uma estratégia para calibração da informação dietética pelo erro de medida.

## MÉTODOS

Utilizou-se o banco de dados do estudo de validação do QFA, desenvolvido por Slater et al,<sup>16</sup> conduzido com 79 adolescentes com idade entre 14 e 18 anos de uma escola pública do Município de São Paulo em 1999. O detalhamento da metodologia adotada foi descrito no citado artigo.

Para esse estudo, foram utilizados os dados de consumo de energia e macronutrientes de todos os indivíduos que completaram pelo menos três recordatórios de 24h (R24h) e um Questionário de Frequência Alimentar para Adolescentes (QFAA).

O consumo de alimentos registrado pelos R24h e pelo QFAA foi transformado em energia e nutrientes, utilizando-se o programa Virtual Nutri, modificado quanto aos valores de composição nutricional dos alimentos e a inclusão de receitas testadas.<sup>6</sup>

Foram incluídos na análise dados do consumo de energia de indivíduos que estivessem entre 500Kcal e 6.000Kcal, conforme proposto por Andrade et al.<sup>1</sup>

Inicialmente foram calculados a média e o desvio-padrão para o total de energia e de macronutrientes da dieta segundo os R24h e o QFA. Testou-se a normalidade da distribuição por meio da aplicação do teste de Kolmogorov-Smirnov. Em seguida, as estimativas foram ajustadas pela energia usando-se o método dos resíduos,<sup>18</sup> com o objetivo de estimar a fração desses nutrientes que não se correlaciona com a ingestão total de energia.

Para a correção de um dado nutriente, deve-se considerar que  $x_i$  indica a verdadeira ingestão habitual do indivíduo  $i$ , e que  $z_{ij}$  é a ingestão estimada pelo *jésimo* R24h ( $j= 1, 2, 3$ ). Considere-se que  $\bar{z}$  seja a média dos valores dos nutrientes estimada pelos três R24h. Para se obter os valores calibrados do QFAA, a média dos valores do R24h ( $\bar{z}$ ) é regredida nos valores do QFAA, tendo por base o modelo clássico dos erros:

$$X = Z + \varepsilon_z \quad (1)$$

Onde  $X$  é o valor verdadeiro,  $Z$  a ingestão obtida pelo R24h, e o  $E(\varepsilon_z) = 0$ . Portanto,  $E(X) = E(Z)$ .

Supondo-se uma relação linear entre os valores obtidos no QFAA e os valores do R24h, utilizou-se o método proposto por Rosner et al,<sup>14</sup> de calibração por regressão linear, para predizer o verdadeiro consumo de energia e nutrientes  $x_p$ , a partir do valor do nutriente obtido pelo QFAA,  $Q_r$ .

$$X = E(Z/Q) = \alpha + \lambda Q_r$$

As estimativas de  $\alpha$  e  $\lambda$  foram obtidas da regressão de  $z_i$  em  $Q_r$ .

Nessa situação, a inclinação da reta da regressão estimada representada pelo  $\lambda$ , é a informação-chave para correção do erro da relação entre o desfecho e a estimativa da ingestão do nutriente no contexto de um estudo epidemiológico.

Assumindo-se a normalidade de  $X$ ,  $Q$  e  $\varepsilon_Q$ , o que implica no modelo linear de erro definido na equação (1), a variância dos valores preditos,  $\text{Var}(X)$ , é estimada como a variância das medidas calibradas do questionário, segundo Kaaks et al.<sup>10</sup>

$$\text{Var}(X) = \text{Var}(E[Z/Q]) = \lambda^2 \text{Var}(Q)$$

## RESULTADOS

Dos 79 adolescentes da amostra, 39 eram do sexo masculino e a média de idade observada foi de 15 anos.

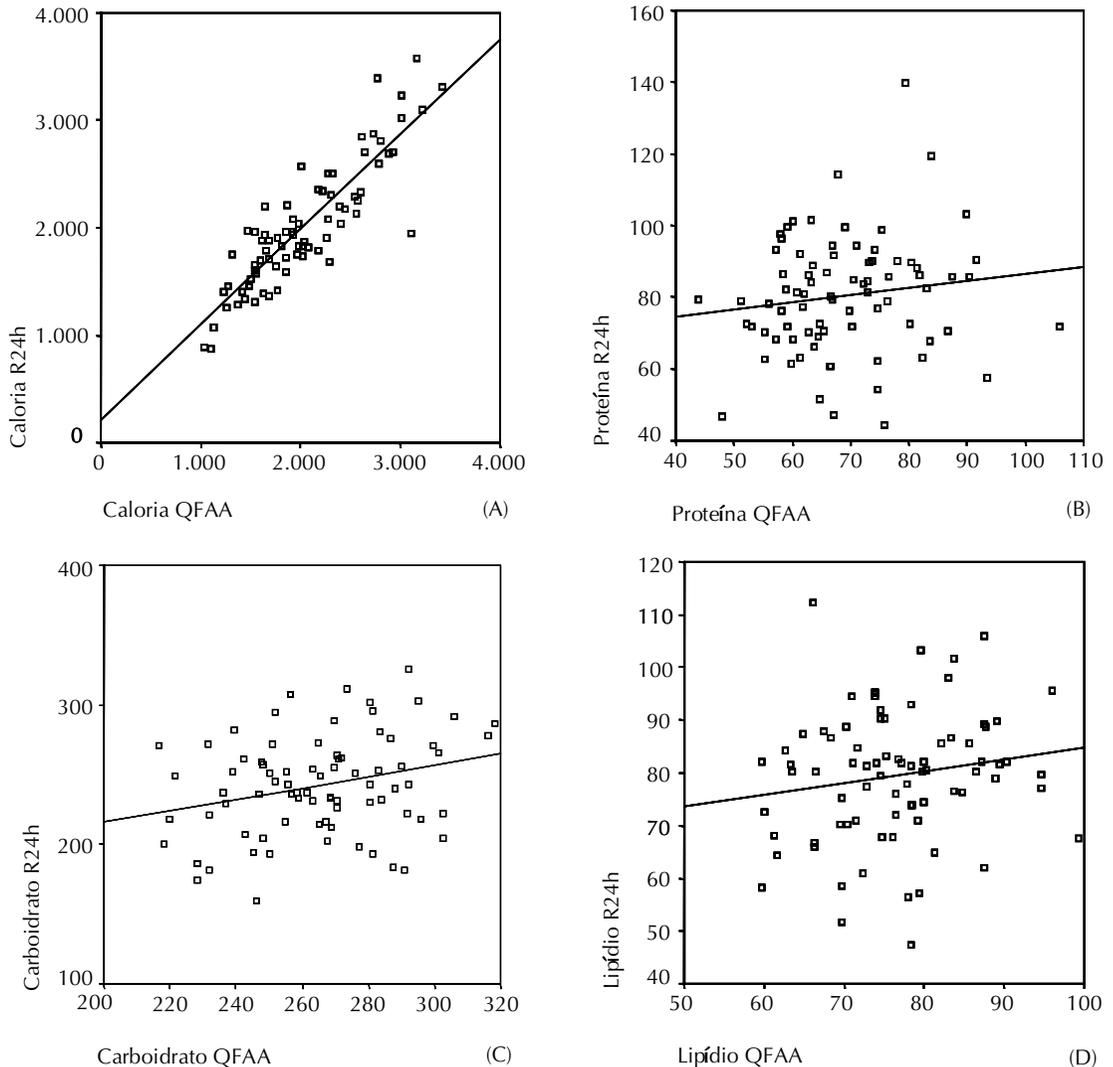
A relação entre os dados dietéticos obtidos pelos dois instrumentos é ilustrada na Figura, representada por gráficos de dispersão para energia e macronutrientes.

Os resultados da aplicação do método de calibração para a correção dos dados de consumo energético e de macronutrientes são apresentados na Tabela 1. As duas primeiras linhas para energia mostram os valores da estatística descritiva da ingestão dietética, medida pelo QFAA e pelo R24h, seguida do valor calibrado. Para os demais nutrientes, são apresentados os valores brutos e ajustados pela energia e, em seguida, os valores calibrados.

Pode-se observar que os valores do R24h, QFAA e valores calibrados foram próximos para energia, com ligeira superestimação pelo QFAA. Também para lipídios, os valores obtidos foram próximos entre os métodos, com valores ligeiramente mais elevados para a estimativa obtida pelo R24h. Para proteínas e carboidratos foram registrados valores pouco mais elevados no R24h e no QFAA respectivamente. Após a calibração, as médias dos QFAA tornaram-se semelhantes às do R24h ajustado pela energia, como desejado (diferenças não significativas no Teste t pareado). No entanto, essa consistência é obtida às expensas da variação. De fato, a variação na ingestão dietética dos valores calibrados diminuiu, como pode ser constatado pelo desvio-padrão da medida calibrada em comparação com os valores originais.

Os valores estimados para os parâmetros  $\alpha$  e  $\lambda$  do modelo de regressão linear, bem como o erro-padrão e o coeficiente de correlação de Pearson, são apresentados na Tabela 2.

Um valor elevado do  $\hat{\lambda}$  significa que aquela particular ingestão do nutriente pode ser medida pelo QFAA de forma comparável ao R24h.



**Figura.** Gráficos de dispersão entre os valores do Recordatório de 24h e os do Questionário de Frequência Alimentar para Adolescentes para energia e macronutrientes ajustados pela energia. Município de São Paulo, 1999. (N=79)

O fator de calibração  $\lambda$ , que idealmente deve ser 1, foi de 0,89 para energia. Para os macronutrientes, os fatores foram de 0,41, 0,22 e 0,20 para carboidratos, lipídios e proteínas, respectivamente.

## DISCUSSÃO

O presente estudo é o primeiro a ser publicado no Brasil apresentando uma das estratégias utilizadas para corrigir os dados dietéticos pelo erro de medida obtidos por QFA. Enfatiza-se que estudos abordando a metodologia de calibração são escassos na literatura, e, apesar das definições claras distinguindo os conceitos de validação e calibração, ainda é comum nos artigos publicados a interpretação ou denominação incorreta de procedimentos e análises estatísticas, especialmente em relação à calibração.

Na calibração por modelo de regressão linear,<sup>14</sup> método utilizado no presente estudo, deseja-se que o intercepto seja aproximadamente zero e a inclinação da reta, representada pelo  $\lambda$ , seja aproximadamente 1. Essas características indicam ausência de vieses no questionário, isto é, a média de ingestão estimada por meio do questionário será igual à média estimada pelo método de referência. Na prática, a inclinação da reta da regressão é menor que 1. Verificou-se um valor de coeficiente de regressão próximo do desejado para a energia ( $\lambda=0,89$ ), indicando um excelente desempenho do instrumento. Estudos semelhantes na literatura relataram valores mais baixos, variando entre 0,09 a 0,45 segundo gênero e etnia.<sup>9,17</sup> Para carboidrato, o coeficiente obtido foi razoável, sendo comparável ao estudo de Stram et al,<sup>17</sup> no Havaí, variando de 0,41 a 0,54 em homens e de 0,20 a 0,73 em mulheres.

**Tabela 1.** Estatística descritiva da ingestão de energia e macronutrientes obtidos por QFAA e R24h de 79 indivíduos, antes e após o ajuste pela energia e a correção. Município de São Paulo, 1999.

Nutriente	Média	Desvio-padrão	IC 95%	Mínimo-máximo
<b>Energia (Kcal)</b>				
QFAA bruto	2.023,6	563,4	1.897,4 ; 2.149,8	1.026,3 - 3.426,5
R24h bruto	2.004,9	570,2	1.877,2 ; 2.132,6	878,2 - 3.578,6
Valor corrigido	2.005,1	498,6	1.893,4 ; 2.116,8	1.122,5 - 3.246,7
<b>Proteína (g)</b>				
QFAA bruto	68,5	20,5	64,7 ; 73,6	20,0 - 123,6
QFAA ajustado	68,9	11,5	66,3 ; 71,5	43,9 - 105,9
R24h bruto	78,9	26,8	73,6 ; 85,5	169,7 - 78,9
R24h ajustado	80,86	16,3	76,7 ; 84,0	44,5 - 139,7
Valor corrigido	80,4	2,3	79,8 ; 80,9	75,4 - 87,7
<b>Carboidrato (g)</b>				
QFAA bruto	263,1	77,9	248,4 ; 282,4	76,0 - 457,1
QFAA ajustado	265,3	23,8	260,0 ; 270,6	216,8 - 318,2
R24h bruto	240,1	74,0	226,0 ; 258,4	72,5 - 407,9
R24h ajustado	242,3	35,3	234,3 ; 250,2	159,9 - 325,5
Valor corrigido	242,2	9,6	239,99 ; 244,30	222,5 - 263,6
<b>Lipídios (g)</b>				
QFAA bruto	76,0	25,2	71,1 ; 82,2	24,8 - 146,7
QFAA ajustado	76,6	9,1	74,5 ; 78,6	59,6 - 99,2
R24h bruto	78,0	28,5	72,3 ; 85,0	23,2 - 170,9
R24h ajustado	79,5	12,4	76,8 ; 82,3	47,3 - 112,2
Valor corrigido	79,5	2,0	79,1 ; 80,0	75,8 - 84,5

QFAA: Questionário de Frequência Alimentar para Adolescentes

R24h: Recordatório de 24 horas

**Tabela 2.** Estimativas dos parâmetros  $\alpha'$  e  $\lambda'$ , erros-padrão e coeficientes de correlação de Pearson ( $r$ ) de energia e macronutrientes ajustados pela energia. Município de São Paulo, 1999.

Nutriente	$\alpha'$	$\lambda'$	Erro-padrão	$r$
Energia	214,20	0,89	0,06	0,87
Proteína	66,64	0,20	0,16	0,14
Carboidratos	134,71	0,41	0,16	0,27
Lipídio total	62,83	0,22	0,15	0,16

Em contrapartida, para lipídios e proteína, considerável viés foi observado (respectivamente  $\lambda=0,22$  e  $\lambda=0,20$ ), denotando substancial correção para aproximação do valor de referência. Os valores de  $\lambda$  obtidos por Kaaks et al<sup>10</sup> para consumo de proteínas em fase piloto do *European Prospective Investigation into Cancer and Nutrition* (EPIC) variaram de 0,23 (mulheres italianas) a 0,43 (homens holandeses). Stram et al<sup>17</sup> relataram valores de  $\lambda$  para ingestão de proteínas para homens (de 0,25 a 0,39) levando-se em consideração a etnia, enquanto que para as mulheres os valores observados variaram de 0,27 a 0,56. Para lipídios, os valores observados variaram entre 0,53 e 0,57 para homens e de 0,34 a 0,68 para mulheres.

No presente estudo, as variáveis foram ajustadas pela energia pelo método dos resíduos, largamente utilizado na literatura, mostrando grande consistência entre a média ajustada e a média original. O ajuste pela energia é motivado tanto pela necessidade de considerar modelos isocalóricos, quanto para controlar o erro de medida embutido nos métodos.<sup>5,11,18</sup> Os coeficientes de correlação de Pearson, após o ajuste, foram elevados para energia e baixos para os macronutrientes, mostrando consistência com as estimativas do fator de calibração.

Uma premissa do método é a normalidade da distribuição das variáveis do modelo.<sup>7</sup>No presente trabalho, as

variáveis apresentaram distribuição normal e os pontos de corte para a exclusão de valores não plausíveis foram estabelecidos.

A atenuação observada no valor da inclinação pode ser explicada parcialmente pelo viés no relato da ingestão, decorrente de erro na estimativa obtida pelo método de referência, devido ao efeito que pode existir em função do tempo e pela diferença na transformação da informação dietética para a ingestão de nutrientes entre os métodos.<sup>4,15</sup>

Outra justificativa relacionada ao efeito de atenuação do coeficiente  $\lambda$  é a violação de pressupostos teóricos do método de calibração, tais como o de independência entre os erros de ambos os métodos de avaliação do consumo alimentar, a ausência de erros sistemáticos nos métodos de referência e a independência entre os erros e a ingestão verdadeira.<sup>18</sup>

Um dos resultados observados foi o encurtamento do desvio-padrão para os dados calibrados. Esse comportamento também foi descrito em outros estudos.<sup>12,15</sup> Os dados calibrados têm menos variação que os dados originais devido à correção dos erros de classificação dos indivíduos. Especialmente, valores extremos são afetados pela correção linear devido ao pressuposto de linearidade entre o método de referência e o questionário.<sup>13</sup>

Fraser e Stram<sup>7</sup> demonstraram o viés presente quando estimativas brutas dos dados do QFA são utilizadas em uma regressão multivariada para estimar o efeito da dieta na doença e que é essencialmente eliminado por uma regressão de calibração. Dessa forma, a utilização do QFA em estudos epidemiológicos sem a correção, poderá incorrer na situação descrita acima. Os autores chamam a atenção para a vasta literatura sobre estudos da relação dieta-doença que, em geral, não utilizam técnicas para correção do erro, o que pode explicar resultados conflituosos ou inconsistentes.

No presente estudo, pôde-se observar que a abordagem de calibração linear aproxima os valores corrigidos à média dos valores do método de referência, indicando uma redução do erro de medida. Porém, a aplicação do método de regressão linear apresentada levanta questões que merecem maior debate. Entre os pontos a serem destacados está o desempenho dessa abordagem face ao não atendimento aos pressupostos teóricos, o que é comum em estudos da dieta que usam como referência métodos de medida baseados no relato dos indivíduos.

Com isso, pode-se ressaltar a necessidade de estudar novas metodologias para correção pelo erro de medida, ou mesmo buscar novas alternativas que tornem o QFA um instrumento menos imperfeito.

## REFERÊNCIAS

1. Andrade RG, Pereira RA, Sichieri R. Consumo alimentar de adolescentes com e sem sobrepeso do município do Rio de Janeiro. *Cad Saúde Pública*. 2003;19:1485-95.
2. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Oxford: Oxford University Press; 1995.
3. Beaton GH. Approaches to analysis of dietary data: relationship between planned analyses and choice of methodology. *Am J Clin Nutr*. 1994;59(1 Suppl): 253S-61.
4. Carroll RJ, Freedman L, Pee D. Design aspects of calibration studies in nutrition with analysis of missing data in linear measurement error models. *Biometrics*. 1997;53:1440-57.
5. Ferrari P, Kaaks R, Fahey MT, Slimani N, Day NE, Pera G et al. Within- and Between-Cohort variation in measured macronutrient intakes, taking account of measurement errors, in the European Prospective Investigation into Cancer and Nutrition Study. *Am J Epidemiol*. 2004;160(8):814-22.
6. Fisberg RM, Slater B. Manual de receitas e medidas caseiras para cálculo de inquéritos alimentares: manual elaborado para auxiliar o processamento de dados de inquéritos alimentares. São Paulo: Signus; 2002.
7. Fraser GE, Stram DO. Regression calibration in studies with correlated variables measured with error. *Am J Epidemiol*. 2001;154:836-44.
8. Fraser GE, Shavlik DJ. Correlations between estimated and true dietary intakes. *Ann Epidemiol*. 2004;14:287-95.
9. Johansson I, Hallmans G, Wikman A, Biessy C, Riboli E, Kaaks R. Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort. *Public Health Nutr*. 2002; 5(3):487-96.
10. Kaaks R, Riboli E, Van Staveren W. Calibration of dietary intake measurements in prospective cohort studies. *Am J Epidemiol*. 1995;142:548-56.
11. Kaaks R, Ferrari P, Ciampi A, Plummer M, Riboli E. Part H. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutr*. 2002;5(6A):969-76.
12. Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, RJ, et al. Part E. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr*. 2002;5(6A):915-23.
13. Kynast-Wolf G, Becker N, Kroke A, Brandstetter BR, Wahrendorf J, Boeing H. Linear regression calibration: theoretical framework and empirical results in EPIC, Germany. *Ann Nutr Metab*. 2002;46:2-8.

14. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med*. 1989;8:1051-69.
15. Sempos CT, Looker AC, Johnson CL, Woteki CE. The importance of within-person variability in estimating prevalence. In: Macdonald I, editor. *Monitoring dietary intakes*. New York: Springer-Verlag; 1991. p. 99-109.
16. Slater B, Philippi ST, Fisberg RM, Latorre MRDO. Validation of a semi-quantitative adolescent food frequency questionnaire applied at a public school in São Paulo, Brazil. *Eur J Clin Nutr*. 2003;57:629-35.
17. Stram DO, Hankin JH, Wilkens LR, Pike MC, Monroe KR, Park S, et al. Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. *Am J Epidemiol*. 2000;151:358-70.
18. Willett WC. *Nutritional epidemiology*. New York: Oxford University Press; 1998. (Monographs in Epidemiology and Biostatistics, 30).