

Sinara L Rossato^{I,II,III}Sandra C Fuchs^{I,II}

Manejo de erros aleatórios e vieses em métodos de avaliação de dieta de curto período

Handling random errors and biases in methods used for short-term dietary assessment

RESUMO

Estudos epidemiológicos têm evidenciado o efeito da dieta na incidência de doenças crônicas, mas a precisão e a acurácia de dados de ingestão alimentar requerem planejamento, delineamento e modelagem estatística. A estimativa da ingestão alimentar usual na população por métodos de avaliação de curto período, como recordatórios alimentares de 24 horas ou diários alimentares, é influenciada por erros aleatórios e vieses inerentes ao método. Para o manejo de erros aleatórios, utilizam-se a modelagem estatística e o apropriado delineamento e amostragem, cruciais para controle de vieses. O objetivo deste artigo é analisar potenciais vieses e erros aleatórios, suas influências nos resultados e como prevenir e/ou tratá-los estatisticamente em estudos epidemiológicos de avaliação de dieta.

DESCRITORES: Registros de Dieta. Análise de Dados, métodos. Ingestão de Alimentos. Consumo de Alimentos. Inquéritos sobre Dietas, métodos.

^I Hospital de Clínicas de Porto Alegre. Porto Alegre, RS, Brasil

^{II} Programa de Pós-Graduação em Epidemiologia. Faculdade de Medicina. Universidade Federal do Rio Grande do Sul. Porto Alegre, RS, Brasil

^{III} Department of Nutrition. Harvard School of Public Health. Boston, Massachusetts, USA

Correspondência | Correspondence:

Harvard School of Public Health
Department of Nutrition – Room 300
677 Huntington Avenue
Boston, Massachusetts 02115
E-mail: sinara.rossato@gmail.com

Recebido: 25/9/2013

Aprovado: 11/3/2014

Artigo disponível em português e inglês em:
www.scielo.br/rsp

ABSTRACT

Epidemiological studies have shown the effect of diet on the incidence of chronic diseases; however, proper planning, designing, and statistical modeling are necessary to obtain precise and accurate food consumption data. Evaluation methods used for short-term assessment of food consumption of a population, such as tracking of food intake over 24h or food diaries, can be affected by random errors or biases inherent to the method. Statistical modeling is used to handle random errors, whereas proper designing and sampling are essential for controlling biases. The present study aimed to analyze potential biases and random errors and determine how they affect the results. We also aimed to identify ways to prevent them and/or to use statistical approaches in epidemiological studies involving dietary assessments.

DESCRIPTORS: Diet Records. Data Analysis, methods. Eating. Food Consumption. Diet Surveys, methods.

INTRODUÇÃO

A estimativa da ingestão de alimentos e nutrientes envolve erros aleatórios e sistemáticos inerentes ao método de coleta dos dados, seja o recordatório alimentar de 24 horas (R24h), seja o diário alimentar (DA). Informações obtidas em um único R24h e DA não representam a ingestão usual. Para que possam representar a ingestão alimentar usual, dependem da colaboração do participante e do número de dias relatados. Ainda assim, médias geradas de várias medidas repetidas têm maior variância e podem resultar em erros de estimativa presentes em uma parcela da população que relata ingestão diferente da verdadeira.² Assim, a avaliação de um único dia ou de vários dias é suscetível a erros, que podem ser minimizados com adequada abordagem estatística e de amostragem.

Quando o erro deriva da variação nas escolhas individuais de alimentos, que naturalmente diferem de um dia para o outro, trata-se de um erro aleatório, comum a todos os indivíduos da população. Porém, além de características individuais, a variabilidade da ingestão alimentar decorre do nível de desenvolvimento do país, local em que o estudo é realizado, características específicas da população e métodos de coleta dos dados. Quando alguma dessas características modifica os resultados, trata-se de um viés e não mais de um erro aleatório.⁶ Nesse sentido, a variação do consumo de alimentos caracteriza um viés quando a ingestão energética é diferente entre verão e inverno, dias da semana e de final de semana e quando indivíduos obesos sub-relatam a ingestão alimentar. Vieses também podem estar relacionados ao desfecho do estudo. Em estudos de caso-controle, os indivíduos incluídos como casos podem recordar a dieta diferentemente daqueles incluídos como controle.³

Tanto o erro aleatório quanto o sistemático têm impacto importante na análise dos dados e interpretação dos resultados.

O objetivo deste artigo é analisar potenciais vieses e erros aleatórios, suas influências nos resultados e como prevenir e/ou tratá-los estatisticamente em estudos epidemiológicos de avaliação de dieta.

Os R24h e DA têm sido largamente utilizados como teste padrão para validação de questionários de frequência de consumo alimentar (QFA), e as estratégias usadas na sua aplicação determinam a acurácia e precisão do método. Reconhecida a variabilidade do consumo alimentar em um indivíduo e a necessidade de utilizar mais que um recordatório para caracterizar a dieta usual, cabe ao pesquisador planejar a amostragem de modo a minimizar potenciais vieses e garantir o poder estatístico do estudo.⁶ Para tal, não apenas o cálculo de tamanho da amostra deve ser realizado, como deve ser determinado o número de medidas repetidas por indivíduo, baseado na razão entre valores das variâncias intra e interindivíduos para nutrientes de interesse.^{1,5} Um dos métodos utilizados no cálculo do número de dias necessários para estimar a ingestão usual baseia-se no nível de correlação entre a ingestão esperada e a usual [$d = [r^2 / (1 - r^2)] \sigma_w / \sigma_b$], onde d é o número de dias de coleta por indivíduo; r é a correlação esperada entre os valores usuais e os observados; e σ_w / σ_b é a razão entre as variâncias intra e interindivíduos. Quanto maior o valor de r , maior será a proporção de indivíduos corretamente classificados, e quanto menor for a razão entre as variâncias, menor será o número de dias necessários para classificar adequadamente os indivíduos.⁵

Outro método baseia-se na determinação do nível de confiança das estimativas de ingestão alimentar, expressas em porcentagem [$d = (Z_{\alpha} CV_w / D_o)^2$], onde d é o número de dias necessários por pessoa que, quando normal, assume o valor de 1,96; CV_w é o coeficiente de variação intraindivíduo calculado dividindo-se a variância intraindivíduo pela média de ingestão; D_o é o limite especificado de erro (nível de confiança) que pode variar de 10,0% a 30,0%.⁵ Caso não tenha sido realizado o cálculo, a interpretação dos resultados não significativos pode ser confirmada estimando-se o poder estatístico, obtido com o número de medidas repetidas realizadas.

Estimativa de tamanho de amostra pode ser feita a partir de estudos realizados em populações semelhantes. Em mulheres adultas japonesas, por exemplo, o número de dias necessários para obtenção de dados fidedignos de ingestão alimentar variou de três a 10 dias utilizando R24h para estimar a ingestão de energia e macronutrientes. O estudo de nutrientes com alta variabilidade de ingestão, como colesterol e vitaminas A e C, pode demandar de 20 a 50 recordatórios. Assumindo-se que o erro na estimativa de ingestão usual varie entre 10,0% e 20,0%, respectivamente, seria necessário investigar 10 e três dias para energia; 91 e 23 dias para colesterol; 118 e 30 dias para zinco.⁷ Basiotis et al¹ estudaram 13 homens e 16 mulheres ao longo de um ano, avaliando a diferença entre o número de dias necessários para avaliar a dieta usual entre grupos, individualmente e para diferentes nutrientes, considerando a precisão estatística esperada. Demonstrou-se que o número de dias necessários para avaliar a ingestão de nutrientes varia de acordo com o nutriente e de pessoa para pessoa. Para avaliar a ingestão de energia foram necessários menos dias do que para avaliar vitamina A, visto que energia é consumida por todos e, embora ambas sejam diferentes entre indivíduos, a variação da energia é substancialmente menor do que a vitamina A (14 para energia entre homens e mulheres e para vitamina A 115 para mulheres e 152 para homens). Para alcançar precisão estatística de 10,0% para cada indivíduo, foi necessário grande número de dias, ao passo que, para a população inteira, o número de medidas repetidas necessárias foi substancialmente menor. Nesse sentido, concluiu-se que o número amostral de indivíduos e de medidas repetidas são cruciais para aumentar a precisão estatística do estudo.¹

INFLUÊNCIA DO ERRO ALEATÓRIO E DA MODELAGEM ESTATÍSTICA

O erro aleatório frequentemente induz a interpretações equivocadas. Dood et al² mostram que o erro aleatório aumenta a amplitude dos resultados, comparando a amplitude de dados de dieta obtidos a partir da coleta de um único R24h com dados de dieta obtidos a partir da coleta de dois ou mais. Usando como

exemplo a ingestão de frutas e verduras, a proporção de indivíduos com ingestão inferior a uma porção diária variou de 9,3% para a ingestão estimada por um único R24h a 0,4%, para a média de dois R24h. O segundo erro comum ocorre na interpretação de testes de hipóteses. A variabilidade excessiva resulta em perda de poder estatístico e é possível invalidar o emprego de testes estatísticos.²

Assumindo-se que os dados de ingestão alimentar estão isentos de vieses, modelagens estatísticas podem atenuar a variabilidade inerente.² O método proposto pelo *National Research Council* (1986) deu origem a, pelo menos, outros seis métodos: método Slob (1993), Wallace (1994), Buck original e modificado (1995), Nusser (2000), Gay (2000) e N-Nusser;⁴ mais recentemente, outros métodos têm sido propostos. A Tabela descreve, passo a passo, diferentes métodos de modelagem estatística para ajustar a variabilidade da ingestão alimentar, a partir da publicada por Dodd et al,² acrescida de informações sobre os modelos *Statistical Program to Assess Dietary Exposure* (SPADE) e *Multiple Source Method* (MSM).

Detalhes sobre o desenvolvimento dos métodos *National Research Council/Institute of Medicine*, *Iowa State University* (ISU), *Best-Power*, *Iowa State University Foods* (ISUF),⁴ MSM e SPADE podem ser obtidos nas referências específicas (Tabela). Outros métodos já foram descritos, adaptados ou remodelados. O método Slob apresentou desvantagens em relação à correção de perdas da variabilidade intraindivíduo, afetando a média nos percentis mais baixos, e o método Buck reproduziu a assimetria dos dados originais.⁴ Assim, foi aperfeiçoado em 2006,⁴ resultando em um software estatístico para estimativa da ingestão usual, o *Age-mode*, readaptado para gerar o software SPADE (Tabela). Diferente de outros modelos, o SPADE descreve a ingestão alimentar em função da idade e apresentou diferença na amplitude dos dados para crianças, em comparação ao método ISU. O método MSM é passível de ser aplicado para estimar o consumo esporádico de alimentos e para QFA e questionários de propensão alimentar. No entanto, também apresentou problemas quando o resíduo de modelos de regressão não tem distribuição normal; por isso, o modelo está sendo aperfeiçoado.

CONSIDERAÇÕES FINAIS

Dados de ingestão alimentar são suscetíveis a erros aleatórios e devem ser submetidos à modelagem estatística, aumentando assim a precisão das estimativas e viabilizando uma adequada interpretação dos resultados. Para a maioria dos estudos, a escolha do método pode não interferir expressivamente nos resultados; entretanto, métodos mais atuais, como o ISUF, MSM e SPADE, podem ser utilizados. Se a avaliação do consumo esporádico de alimentos ou nutrientes é o foco de interesse, o método MSM é preferível e brevemente deverá estar

Tabela. Modelos estatísticos para derivação de ingestão alimentar usual baseada em R24h e DA.

NCR/IOM	ISU	BP	ISUF	MSM	SPADE
Passo 0: Aplicação do ajuste inicial					
Aplica <i>Power</i> ou <i>log</i> transformação nos dados dos R24h observados até que a distribuição dos dados esteja próxima da normal.	Ajusta os R24h observados para vieses não individuais, como estações do ano, dias da semana e efeito de amostragem.	Aplica <i>Power</i> ou <i>log</i> transformação nos dados dos R24h observados até que a distribuição dos dados esteja próxima da normal.	Estima a distribuição da probabilidade de consumo de um dado dia, baseado na frequência relativa de valores diferentes de zero nos R24h.	Aplica transformação Box-Cox para obter uma distribuição próxima da normal.	Aplica transformação Box-Cox para obter uma distribuição próxima da normal.
	Constrói uma transformação em dois estágios para aproximar a distribuição dos R24h modificados de uma distribuição normal.	Coloca de lado os zeros dos R24h e ajusta o consumo para vieses não individuais, como estações do ano, dias da semana e efeito de amostragem.			
		Constrói uma transformação em dois estágios para aproximar a distribuição dos R24h transformados de uma distribuição normal.			
Passo 1: Descrição da relação entre R24h individuais medidos e a ingestão individual usual					
Os dados do R24h estimam sem viés a ingestão usual transformada (pressuposto A).	Os dados dos R24h estimam sem viés a ingestão usual na escala não transformada (pressuposto B).	Os dados dos R24h estimam sem viés a ingestão usual na escala não transformada (pressuposto B).	A ingestão usual é a probabilidade de consumo em um determinado dia multiplicado pela ingestão usual total para um dia de consumo alimentar.	Estima a probabilidade de consumo usando regressão logística e a ingestão total diária usando regressão linear.	Monta um modelo polinomial fracionado para os dados transformados.
			Um R24h mede o consumo exatamente igual a zero.		
			Os dados dos R24h com valores diferentes de zero estimam sem viés a ingestão usual na escala não transformada (pressuposto B).		

Continua

Continuação

Passo 2: Separação da variação total das medidas do R24h em variação intra e interindividuais

A variância intraindividual é a mesma para todos os indivíduos.	A variância intraindividual pode variar entre os indivíduos.	A variância intraindividual pode variar entre os indivíduos.	O resíduo transformado é utilizado para estimar a variância inter e intraindividual, a qual é utilizada para reduzir a média alimentar de um indivíduo a uma média geral.	Obtém uma estimativa polinomial fracionada de efeito misto para separar a variabilidade intra da interindividuais em função da idade.
---	--	--	---	---

Passo 3: Estimação da distribuição da ingestão usual contabilizando a variância intraindividual

Constrói um conjunto de valores intermediários que retém a variância entre indivíduos dos dados transformados do R24h.	Constrói um conjunto de valores intermediários que retém a variância entre indivíduos dos dados transformados do R24h.	Transformação inversa: aplica a função inversa da transformação em dois estágios, simultâneos a ajustes para viés; descreve matematicamente a distribuição em escala original da ingestão usual do consumo diário.	Transformação inversa: solução integral dos valores inteiros não negativos dos parâmetros Box-Cox.	Identifica valores discrepantes pelo método Grubbs. Testa a normalidade residual e a distribuição dos dados por Kolmogorov-Smirnov usando o módulo estatístico S-plus. Checa a distribuição de λ .
Transformação inversa: aplica a função inversa para cada valor intermediário.	Transformação inversa: aplica a função inversa para cada valor intermediário.	Transformação inversa: usando a função inversa da transformação inicial <i>Power</i> ou <i>log</i> simultaneamente com ajustes para viés, adequa cada valor intermediário em escala normal para obter a escala original em contrapartida.	A estimativa da ingestão usual é obtida multiplicando a probabilidade de consumo e a ingestão diária total estimada por modelos de regressão.	Quando encontrados valores discrepantes, estes são eliminados e os passos anteriores são repetidos.
A distribuição empírica inversamente transformada é a distribuição da ingestão usual.	A distribuição empírica inversamente transformada é a distribuição da ingestão usual.	A distribuição empírica inversamente transformada é a distribuição da ingestão usual.	Transformação inversa: aplica a transformação inversa com uma função quadrática gaussiana (Monte Carlo Simulations).	

Fonte: adaptado de Dodd et al.² 2006.

R24h: recordatório alimentar de 24h; DA: diário alimentar; NRC: National Research Council; IOM: Institute of Medicine; ISU: Iowa State University; BP: Best-Power; ISUF: Iowa State University foods. Descrição adicional de métodos – MSM: Multiple Source Method;¹⁸ SPAD: Statistical Program to Assess Dietary Exposure⁹

disponível já aperfeiçoado. Vieses podem ser minimizados com seleção amostral e delineamento do estudo apropriados. Para evitar erros sistemáticos relacionados à amostragem, recomenda-se que características como estado nutricional e de saúde, os dias da semana e as estações do ano sejam proporcionais e heterogêneas. O número de medidas repetidas de R24h e o tamanho da amostra podem ser estimados com base na variabilidade da ingestão de nutrientes entre indivíduos. Nutrientes presentes na grande maioria dos alimentos,

como macronutrientes, requerem menor número de medidas repetidas, uma vez que a variabilidade entre as observações é menor. Se o interesse do estudo é avaliar a ingestão alimentar da população de modo geral, amostras maiores, com menor número de medidas repetidas, podem gerar dados com confiabilidade aceitável. No entanto, em estudos de validação, cuja variabilidade individual é crucial, pois será a medida de referência para avaliar a validade, maior número de medidas repetidas é preferível.

REFERÊNCIAS

1. Basiotis PP, Thomas RG, Kelsay JL, Mertz W. Sources of variation in energy intake by men and women as determined from one year's daily dietary records. *Am J Clin Nutr.* 1989;50(3):448-53.
2. Dodd KW, Guenther PM, Freedman LS, Subar AF, Kipnis V, Midthune D, et al. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *J Am Diet Assoc.* 2006;106(10):1640-50. DOI:10.1016/j.jada.2006.07.011
3. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional studies. *J Natl Cancer Inst.* 2011;103(14):1086-92. DOI:10.1093/jnci/djr189
4. Hoffmann K, Boeing H, Dufour A, Volatier JL, Telman J, Virtanen M, et al. Estimating the distribution of usual dietary intake by short-term measurements. *Eur J Clin Nutr.* 2002;56(Suppl 2):S53-62. DOI:10.1038/sj/ejcn/1601429
5. Nelson M, Black AE, Morris JA, Cole TJ. Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *Am J Clin Nutr.* 1989;50(1):155-67.
6. Willett WC. *Nutrition epidemiology.* 3.ed. New York: Oxford University Press; 2013.
7. Tokudome Y, Imaeda N, Nagaya T, Ikeda M, Fujiwara N, Sato J, Kuriki K, Kikuchi S, Maki S, Tokudome S. Daily, weekly, seasonal, within- and between-individual variation in nutrient intake according to four seasons consecutive 7 day nutrient diet records in Japanese female dietitians. *J Epidemiol.* 2002;12:85-92.
8. Department of Epidemiology of the German Institute of Human Nutrition Postdam-Rehbrücke, Versão 1.0.1. Disponível em: <https://nugo.dife.de/msm>
9. Waijers PMCM et al. The potential of AGE_MODE, an age-dependent model, to estimate usual intake and prevalence of inadequate intakes in a population. *J Nutr.* 2006;136:2916-20.

Estudo financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – Processo 70/2008, bolsa de doutorado de Rossato SL), e pelo Hospital de Clínicas de Porto Alegre através do Fundo de Incentivo à Pesquisa e Eventos (FIPE-HCPA – Processo 00-176).

Os autores declaram não haver conflito de interesses.