CSP CADERNOS DE SAÚDE PÚBLICA
REPORTS IN PUBLIC HEALTH

QUESTÕES METODOLÓGICAS
METHODOLOGICAL ISSUES

Correction of diet-outcome association for day-to-day variance in dietaray intake: performance evaluation by simulation

Correção de medidas de associação pela variação do dia a dia no consumo alimentar: avaliação do desempenho por meio de simulação

Corrección de las medidas de asociación por la variación diaria en la ingesta de dietética: evaluación del desempeño por simulación

Eliseu Verly-Jr [1]
Rosely Sichieri [1]
Valéria Troncoso Baltar [2]

doi: 10.1590/0102-311X00173216

Abstract

The use of one or a few 24-hour recalls (24hR) to measure dietary exposure in models to estimate diet-outcome association leads to attenuation of the measure of association and a decrease in the test's power. This is due to daily variation in food intake. The measure of association can be corrected using regression calibration and requires at least one repetition of the 24hR in a subsample of the study population. However, the calibration's performance and the precision of the corrected coefficients can vary according to the characteristics of the study to which it is applied. The aim of this study is to evaluate the performance of correction in different research scenarios in relation to the estimated measure of association and its precision. A population (n = 1,000) was simulated with information on food intake for 200 days and an outcome with an association defined with usual intake (mean for the 200 intake days). The scenarios evaluated were: (a) 100%, 60%, 40%, and 20% of the sample with 2 intake days; (b) individuals with 2, 3, 4, and 5 24hR; and (c) populations with 1,000, 600, and 300 individuals. The coefficients were estimated for 300 random combinations of intake days; mean corrected coefficients were similar to the true coefficient. Precision was lower in all the scenarios: the probability of finding a significant association (when true) varied from 0.47 to 0.29 (100% to 20% with repetition, respectively); 0.47 to 0.78 (2 to 5 days); and 0.47 to 0.15 (1,000 to 300 individuals).

Diet Surveys; Food Consumption; Regression Analysis; Public Health Nutrition

Correspondence
E. Verly-Jr
Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro.
Rua São Francisco Xavier 524, Rio de Janeiro, RJ 20550-013, Brasil.
eliseujunior@gmail.com

[1] *Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.*
[2] *Instituto de Saúde Coletiva, Universidade Federal Fluminense, Niterói, Brasil.*

Introduction

Studies in nutritional epidemiology frequently aim to describe diet-outcome associations. To estimate the measure of association that describes the relationship between diet and outcome requires knowing each individual's disease status and usual dietary intake. Traditionally, usual intake in large cohorts has been collected with the food frequency questionnaire [1]. However, some authors have demonstrated this instrument's high inherent measurement error [2], even identified as one of the reasons for failing to find associations (whether protective or risk) between diet and health outcomes [3,4]. This has led researchers to collect dietary intake using short-term instruments such as 24-hour recalls (24hR) [5,6]. However, due to operational and cost constraints, the collection of more than one 24hR in large epidemiological studies is rarely done, which makes the information on individual intake (for one day or the mean of a few days of 24hR) biased in relation to usual intake (the mean of a large number of 24hR), as a function of daily variation in intake, which is described as intrapersonal variance [7,8]. Since people consume larger or smaller amounts than their usual intake throughout the days, the differences between intake on a given day and the usual intake are considered a random error, since it is distributed randomly around true intake [9]. The effects of random error when describing the relationship between diet and outcome are the attenuation of the measure of association and a decrease in the test's power to detect this association, which can compromise the study's validity [10].

One way of dealing with this error is applying the regression calibration methodology, which estimates the measure of association using each individual's predicted intake as the dietary exposure variable [11]. Kipnis et al. [12] proposed an extension of the calibration method to correct measure of association by day-to-day variation in dietary intake, in addition to accommodating asymmetric and zero-inflated distribution, as in the case of foods that are not consumed every day. Its application requires repetition of the 24hR on non-consecutive days in at least a subsample of the study population. Since it is in part a predictive model for individual intake values, it is important to know how close the estimate is to the real measurement (the one without random error). Another important issue in the calibration's performance is the coefficients' precision, varying according to the study's characteristics, such as sample size, number of repetitions of the 24hR [13], percentage of the study population in which the 24hR is repeated [14], and the variables included in the predictive model [12].

The current study aims to assess the performance of regression calibration for correcting measures of association, in different scenarios, through a simulation study. The simulation is based on a previous study conducted by the authors that collected 20 days of 24hR in a sample of 302 persons. The study provides the necessary parameters for the simulation, allowing the generation of populations with the desired size and data collection days.

Methods

Regression calibration

The following annotations were used to be consistent with those in the international literature: for individual i on day j, $i = 1,..., n$; $j = 1,..., k$; R_{ij} represents 24hR intake (R: reported intake); T_i represents the individual's usual intake i (T: true intake; unbiased intake measured for a long period of time); and Y_i an outcome associated with T_i. A hypothetical association between Y_i and T_i can be described by the following linear regression model:

$$E(Y_i|T_i, Z_i) = m(\beta_0 T_i + \beta^t{}_Z Z_i) \quad (1)$$

In which $Z_i = (Z,...,Z_{ip})$ is a vector with the covariates, measured without errors, for each individual i, and m^{-1} is the link function (in this study, identity). Since usual intake is generally not known (T_i), the model uses the mean from few 24hR for each individual i (R_i) as the dietary exposure variable for obtaining the measure of association between the food and the outcome Y_i,, which leads to a biased (attenuated) estimate of β_T. Calibration consists of the prediction of usual individual intake, based on a two-part mixed effects model that uses the intake obtained from the 24hR as the dependent variable and the same set of variables used for adjusting the diet/outcome model (1) as the independent variables. Predicted individual intakes then sibstitute R_i to obtain a deattenuated estimate of the measure of association between the food and the outcome Y_i. The complete description of the model can be found in Kipnis et al. [12].

This study applies regression calibration to estimate corrected linear regression coefficients for the relationship between R_i (with information on intake from two or more 24HR for each individual) and Y_i obtained by the model (1). Corrected coefficients will be compared to the real coefficients (β_T, with information on intake from 200 24hR for each individual) for each study scenario (described later).

Considering that prediction of intake is a function of the amount of food reported on the collection days, i.e., prediction using the first and second days will be different from prediction using the first and third days, and so on, therefore 300 combinations of two or more 24hR were selected per individual. For each combination, we performed regression calibration to obtain the corrected measure of association and its confidence interval; these were compared to β_T (set at 1.0 in the simulation).

The regression calibration was performed with the mixtran and indivint macros available for SAS (SAS Inst., Cary, USA).

Data simulation

- **Individual intake**

A population of 1,000 individuals was simulated with information on intake for 200 days for each individual. The simulation assumed that T_i (the individual's usual intake *i*) is the product of the mean amount consumed on the intake days (A_i) and the probability of the individual consuming the food (P_i). The amount consumed for each intake day was generated by the equation:

$$T_i = P_i \cdot A_i \quad \} R_{ij} > 0$$
$$R_{ij} = A_i + \varepsilon_{ij} \quad \} R_{ij} > 0 \qquad (2)$$

Where $A_i \sim$ Normal (μ, σ_b^2), $P_i \sim$ Bernoulli (p_i), R_{ij}, represents the individual's intake *i* on day *j*, and $\varepsilon_{ij} \sim$ Normal $(0, \sigma_w^2)$; σ_b^2 and σ_w^2 are the inter and intrapersonal variances, respectively. The random variables (A_i, p_i) were generated with bivariate distribution with correlation ρ_{Ap} between them. Since A_i on the original scale is generally right-skewed, the parameters μ and σ_b^2 and in the normal distribution were generated on the Box-Cox scale, chosen because it is the most widely used in food intake studies with skewed data, and subsequently back-transformed to the original scale. In order to define the days with and without intake, a random variable with Bernoulli distribution was generated with the intake probability defined for each individual (p_i), since the probability of consuming the food varies between persons according to the observed distribution of probabilities in the population. The model considered right-skewed distribution, since it is the one most frequently observed in real data. The correlation between amount consumed and probability of intake was considered in the simulation. The parameters ($\mu, \sigma_b^2, \sigma_w^2$, p_i and ρ_{Ap}) and the distributions and lambda from the Box-Cox transformation were obtained from data collected in the baseline study. The mean of the 200 intake days was calculated for each individual, and was considered as individual usual intakes (T_i).

- **Covariates**

For this population, we generated age values (in years), assuming normal distribution (mean = 25; standard deviation – SD = 5), with a correlation of 0.3 with usual intake. Sex distribution was defined as 50% for each sex, in with mean usual intake and mean age were 20g and 2 years higher for men (sex and age were assumed to be error-free).

- **Outcome**

Next, outcome (Y_i) was simulated, whereby its relationship with usual intake (T_i) was specified by the following linear regression model:

$$Y_i = \beta_0 + \beta_T T_i + \beta_Z Z + \varepsilon_i \quad (3)$$

Where: Y_i the simulated outcome, with normal distribution with mean and SD of 25 and 3, respectively, arbitrarily chosen; β_0 is the intercept; coefficient β_T was set at 1.0 for the relationship with usual intake on the original scale with sample power set at 80% to detect $\beta_T \neq 0$; β_z; is the vector with the effects of the covariates in Z_i: 1 and 5 for sex and age, respectively; ε_i Normal (0, 1). Data were analyzed in the Stata v.13 statistical package (StataCorp LP, College Station, USA).

Scenarios assessed

The following scenarios were tested:
a) Different percentages of the study population answering a second 24hR: 100%, 60%, 40%, and 20%. In each of the 300 combinations of intake days, we selected an intake day for the entire population and a second intake day only for the previously defined percentages.
b) Different numbers of 24hR for each individual in the study population: j = 2, 3, 4, and 5. In each of the 300 combinations of intake days, specific numbers of 24hR were selected for each individual. We also tested a scenario in which different percentages of the population answer different numbers of 24hR: 40% with four 24HR, 30% with three 24hR, and 20% with two 24hR.
c) Different population sizes: 1,000, 600, 300. This item also included a scenario with the necessary sample size calculated to obtain coefficients statistically different from zero: n = 2,400. This size was obtained by simulation, with the lowest value that guaranteed that at least 2.5% of the coefficients were different from zero.

Finally we compared corrected and uncorrected coefficients and their confidence intervals for the scenario with n = 1,000 and 100% of the sample with the second 24hR.

Simulation parameters

All the parameters used in the food intake simulation were obtained from a longitudinal study with 302 participants in the city of Rio de Janeiro, Brazil, in which each answered 20 non-consecutive 24hR. The study used a snowball sampling strategy in which the interviewers were selected (23 undergraduate nutrition students) and they later chose the interviewees. To guarantee adherence to the data collection, the interviewees were preferably from the same social circle or lived close to the interviewers, in addition to expressing their willingness to remain in the study and provide detailed information 20 times on their food intake. Although the sample was not random, the participants were dispersed all across the city. Data were collected from March 2013 to April 2014, with a mean follow-up lenght for each individual of three months. The multiple pass method was used to collect the 24hR intake data [15]. During the fieldwork, the interviewers took the first recalls applied to each participant for the initial data check. The reported foods were entered into the Brasil Nutri program, which is based on foods, serving sizes, and preparations, reported in a national food acquisition survey.

The study was approved by the Institutional Review Board of the Institute of Social Medicine, State University of Rio de Janeiro.

Results

In the simulated population, mean usual intake was equal to the mean one-day intake (78g). SD were 75 and 136 for usual and one-day intake, respectively. Distribution of usual intake was right-skewed (skewness = 1.39, kurtosis = 5.35) with 9.4% of the population consisting of usual non-consumers (intake equal to zero in the mean of the 200 intake days) (Figure 1). Distribution of the one-day intake showed a mean of 60% of non-consumers and was more skewed than the distribution of usual intake (skewness = 2.64, kurtosis = 13.9). Correlation between the outcome and usual intake was 0.30. Mean age was 20.2 year (SD = 4.9), and 51% of the population were males. The simulated outcome showed normal distribution with a mean of 25 (SD = 4.1). The adjusted linear regression coefficient for usual intake was 0.99 (95% confidence interval – 95%CI: 0.32-1.66); for sex it was 0.69 (95%CI: 0.30-1.07); and for age it was 0.52 (95%CI: 0.48-0.56). The full model coefficient of determination was 0.47.

Figures 2, 3 and 4 show the coefficients, corrected for intrapersonal variance, for the association between food intake and outcome, adjusted by sex and age. The solid line represents the corrected coefficients, ordered from lowest to highest, for 300 random combinations of two or more 24hR for each individual, and the gray area represents the respective confidence intervals. The mean of the corrected coefficients in each scenario varied from 0.98 to 1.01. Figure 2 shows the analyses for different percentages of the sample with the second recall. There was an increase in the coefficients' dispersion, reflecting wider confidence intervals as the percentage of the sample with replication of the 24hR decreases from 100% to 20%. Among the 300 combinations of two days of collection, the lower limit of the 95%CI was greater than zero in 47%, 37%, 37%, and 29% for 100%, 60%, 40, and 20% replication, respectively. Likewise, precision increased with the increasing number of repetitions per individual. For these scenarios, the percentage of coefficients whose lower limit of the 95%CI was greater than zero, among the selected combinations, was 47%, 53%, 67%, and 78% for 2, 3, 4, and 5 days of 24hR for each individual, respectively (Figure 3). When testing the scenario with 40% with four days of 24hR, 40% with three, and 20% with two, the percentage of significant coefficients was 61% (data not shown).

Figure 4 shows the coefficients for different sample sizes. The coefficients' precision decreased greatly with the reduction in the sample size. With the sample of 2,400, all the coefficients were statistically different from zero; for scenarios with sample sizes of 600 and 300, the proportion of coefficients whose lower limit of the 95%CI was greater than zero, among the selected combinations, was 25% and 15%, respectively.

Figure 1

Distribution of simulated usual intake.

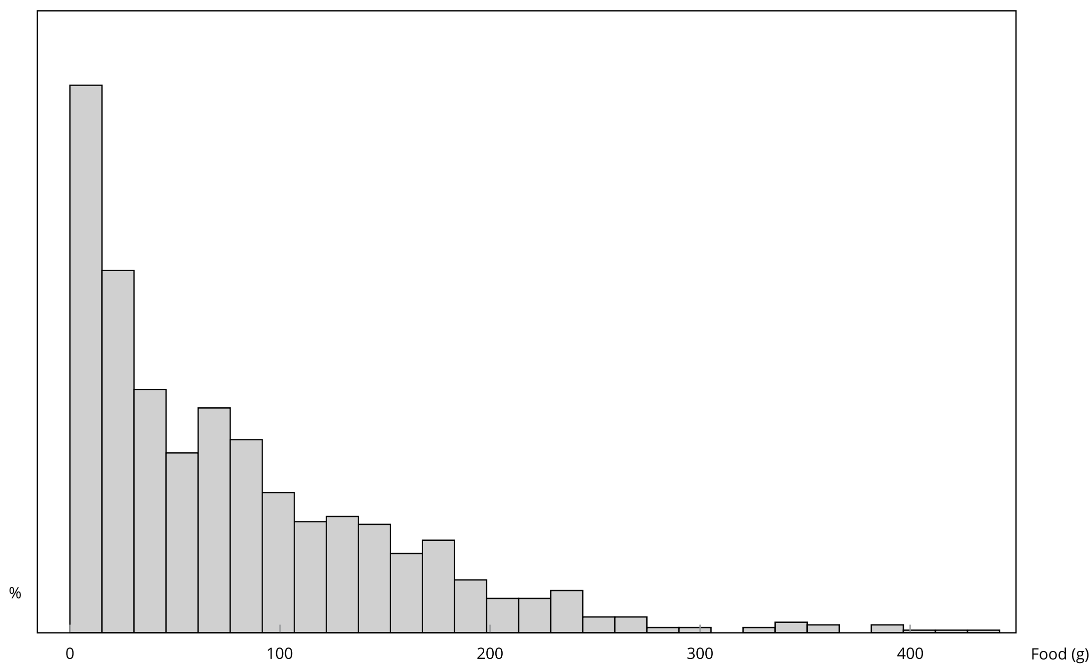
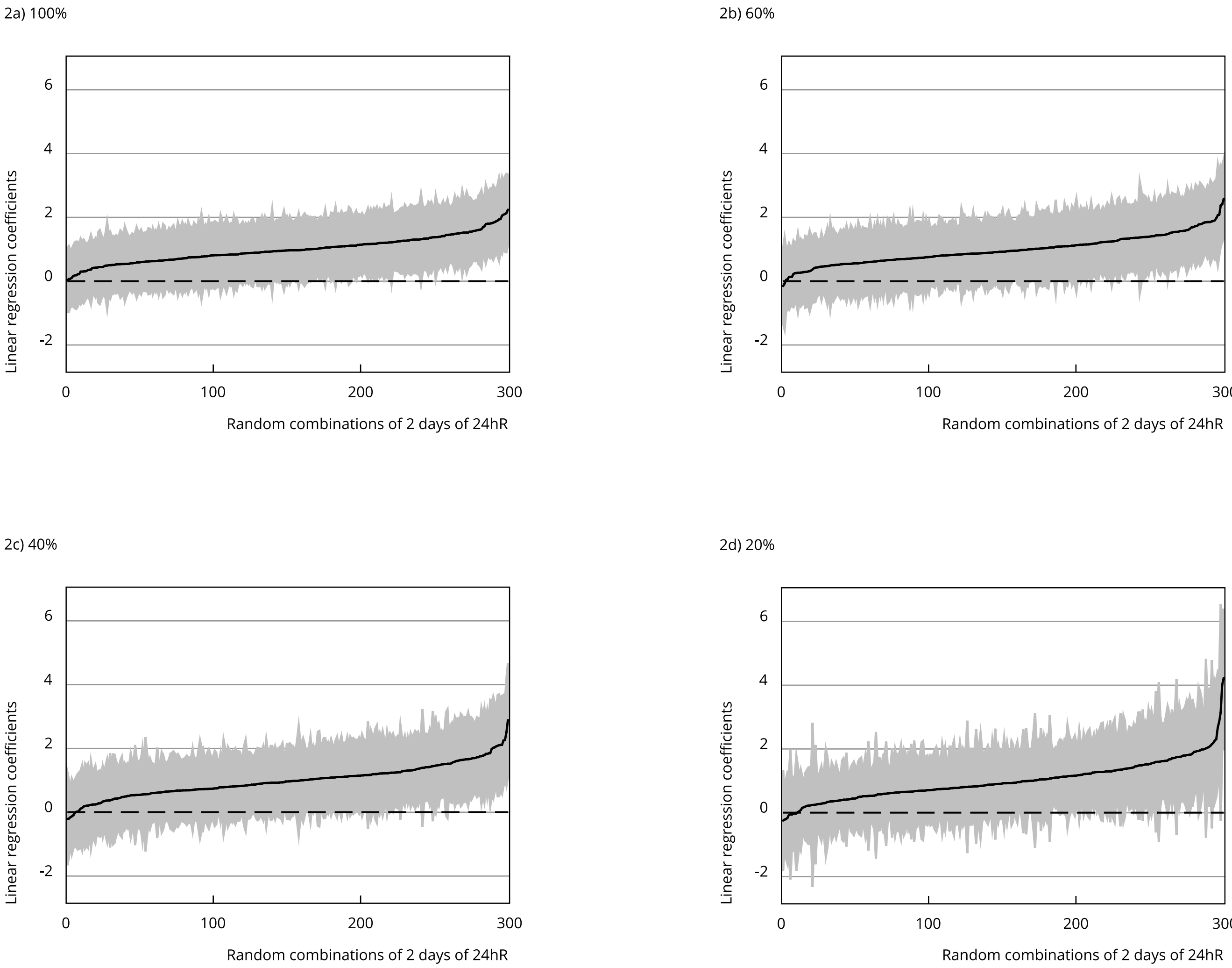


Figure 2

Corrected linear regression coefficients and 95% confidence intervals (95%CI) for the study population with 100%, 60%, 40%, and 20% with repetition of the 24-hour recalls (24hR).



Note: the solid line and gray area represent the corrected coefficients and their confidence intervals, respectively. The coefficients describe the association between food intake and the outcome, adjusted for sex and age.

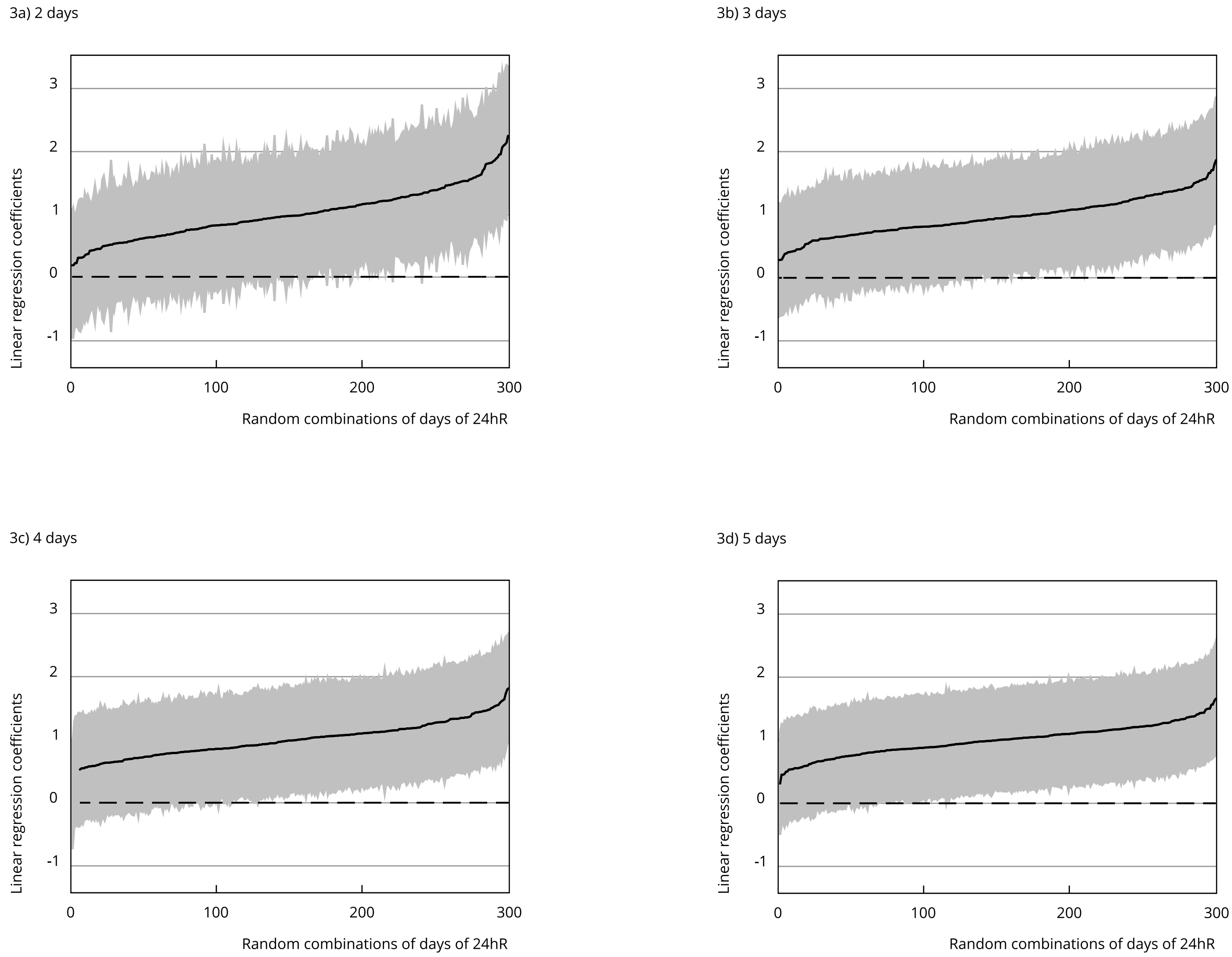
Figure 5 compares corrected and uncorrected coefficients. The mean of the uncorrected coefficients was 0.42. The proportion of coefficients whose lower limit of confidence interval was greater than zero was similar to that of the corrected coefficients (47%).

Discussion

The purpose of calibration is to correct measures of association attenuated by random error, which occurs when few days of 24hR data are collected for each individual in the study population [12]. It is thus expected that the corrected measure of association will be as close as possible to the true measure

Figure 3

Corrected linear regression coefficients and 95% confidence intervals (95%CI) for the study population with 2, 3, 4, and 5 24-hour recalls (24hR) for each individual.



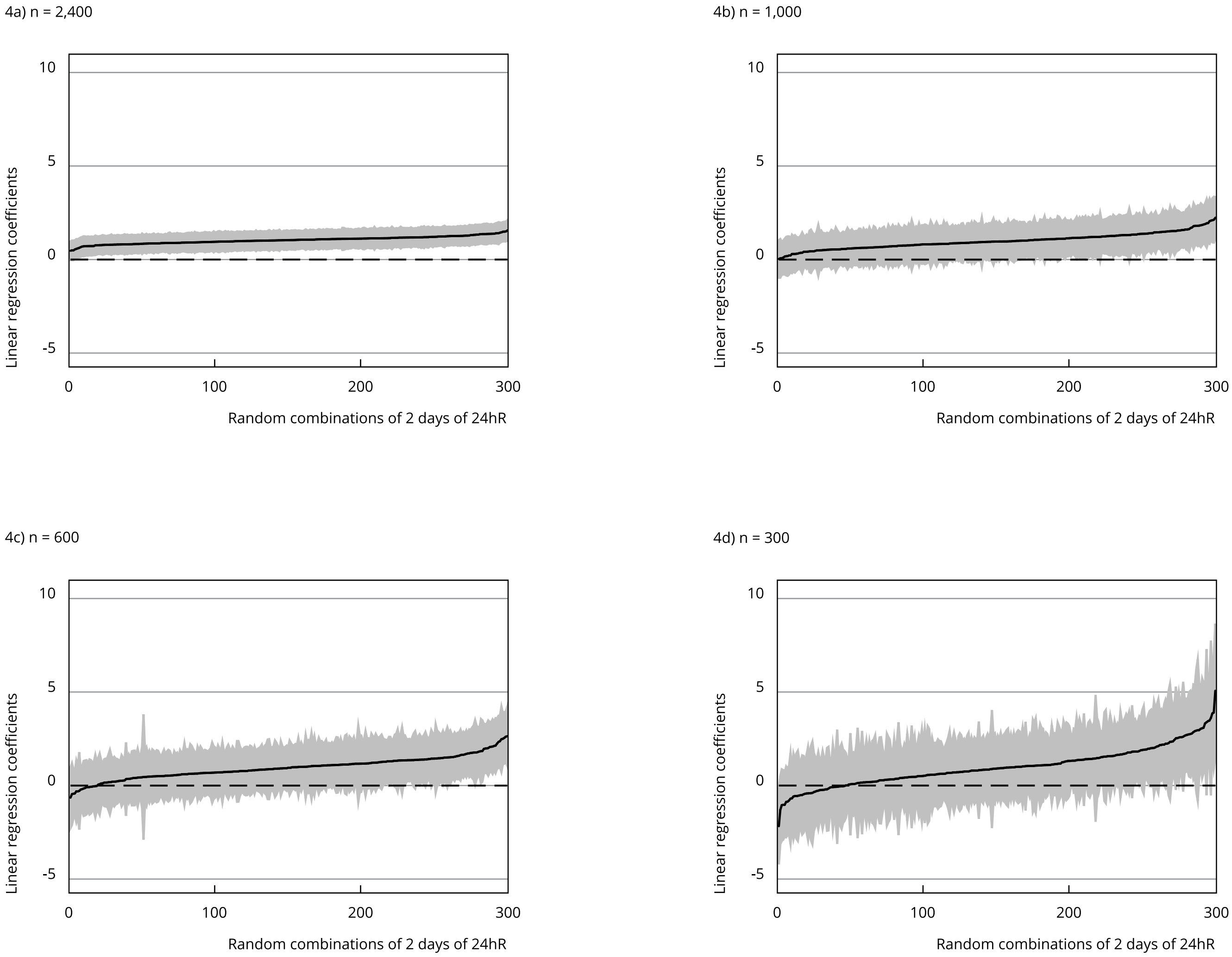
Note: the solid line and gray area represent the corrected coefficients and their confidence intervals, respectively. The coefficients describe the association between food intake and the outcome, adjusted for sex and age.

of association, i.e., the measure that would be obtained if each individual's usual intake were known. However, since different combinations of recall days can generate different measures of association, and thus different corrected coefficients, comparison of the corrected and true measures of association should take into account a large number of possibilities of combinations of recall days.

In this sense, the mean of the coefficients obtained from the combinations of two days of recall should be close to the real coefficient; higher or lower means indicate a tendency towards under or overestimation of the corrected coefficients. Meanwhile, the coefficients' dispersion indicates their precision. The mean coefficients were very similar for all the tested scenarios, with very narrow variation in relation to the real coefficient. However, precision varied according to the scenario, such that the sample's power becomes insufficient to detect an association, even when it really exists and

Figure 4

Corrected linear regression coefficients and 95% confidence intervals (95%CI) for populations with 2,400, 1,000, 600, and 300 individuals.



24hR: 24-h recalls.

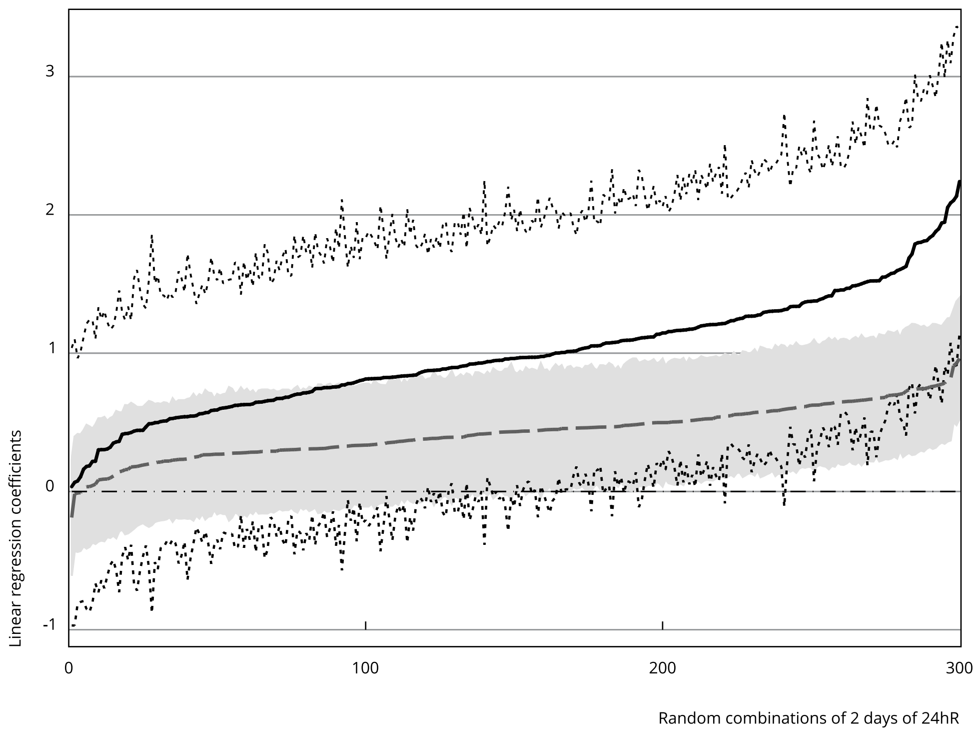
Note: the solid line and gray area represent the corrected coefficients and their confidence intervals, respectively. The coefficients describe the association between food intake and the outcome, adjusted for sex and age.

the sample size has been calculated to be able to detect it. One consequence of the random error is a decrease in the sample's power; this loss of power was not restored after correction, since the percentage of significant coefficients was similar in the corrected and uncorrected analysis. For some scenarios the precision becomes even smaller. When up to two 24hR are used for each individual, the precision decreases significantly when the proportion of individuals with the second 24hR is less than 40%. A previous study had already suggested that between 60% and 40% of repetition is sufficient to maintain precision in estimating usual food intake percentiles [14].

An important issue is sample size. This study simulated an outcome whose association with usual intake (mean of 200 intake days) could be statistically significant with 80% power and sample size of 1,000. Even using n = 1,000, the association was not significant in 53% of the 300 selected combinations of 24hR days. Using the same outcome in smaller samples, especially less than 300, the likelihood

Figure 5

Corrected and uncorrected linear regression coefficients and 95% confidence intervals (95%CI) for the population with 1,000 individuals with 2 24-hour recalls (24hR) for all the individuals.



Note: the solid line and dotted black lines represent the corrected coefficients and their confidence intervals. The dashed line and gray area represent the uncorrected coefficients and their confidence intervals. The coefficients describe the association between food intake and the outcome, adjusted for sex and age.

of finding an association is decreased substantially, even when it really exists. The extent to which a reduction in sample size will increase the probability of type error 2 (not rejecting the null hypothesis when it is false) will depend on the true coefficient's effect size and precision [16]. In a real scenario, in which the effect of diet on health outcomes are usually small [17,18], any loss of precision may decrease the probability of rejecting the null hypothesis, even if it is false. In this simulation, while n = 1,000 would be sufficient to find an association with usual intake, drawing on correction with two or more intake days, the estimated necessary size was 2,400.

Another way of increasing the coefficients' precision is to increase the number of 24hR repetitions for each individual, which was observed in this study when increasing from two to five days of 24hR. Carroll et al. [13], using real and simulated data, found that between four and six days of 24hR for each individual would be sufficient for the majority of the dietary items. Therefore, correction for random error requires a sufficient sample size or number of repetitions to obtain measures of association that represent the real association. Considering the difficulty in obtaining various days of 24hR in large epidemiological studies, one possibility is to collect more repetitions in a subsample. In this simulation, similar results were observed in scenarios in which all the individuals answered four 24hR and in which a subsample answered four, three, and two 24hR.

Importantly, these results refer to a sample of 1,000 individuals with 80% power to detect an association (β = 1) between usual intake and the outcome in a multiple model. Thus, it is not a general recommendation for planning new studies; adequate sample size will depend on each study's objectives, including food variance and the target outcome, the expected effect size, and the covariates in the predictive model [19]. In addition, the decision to increase the sample or the number of repetitions should be based on the costs involved in each procedure. The researcher should assess whether the increase in cost and fieldwork time compensates for the gain in precision. A simulation study can assist planning new studies with an estimate of the best combination of sample size and 24hR repetitions, so as to optimize efficiency in data analysis.

Both Carroll et al. [13] and Kipnis et al. [12] found substantial improvement in the prediction of some items by including intake frequency as a variable in the prediction model; other variables related to food intake such as socioeconomic variables, body mass index, and others, even those not present in the diet/outcome model (equation 3 in the methods section) can be included in the predictive model for usual intake and potentially increase the corrected coefficients' precision.

Importantly, the method only proposes correction for random error; the effects of other types of errors such as underreporting or differential error are not reduced. The latter is particularly important in cross-sectional studies and some types of case-control studies in which disease status can interfere in intake report and modify the direction of the measure of association (reverse causality) [20], which is not resolved with regression calibration. Finally, the study tested the identity link function by providing a direct interpretation of the relationship between dietary intake and the outcome. An example of the method's application involves estimating the degree to which blood pressure increases in mmHg for each 1,000mg of sodium consumed. However, the method can be applied to other link functions such as log or logit function [12].

In conclusion, correction for random error will produce coefficients close to the true coefficient as long as the sample size or number of repetitions per individual is sufficient to guarantee the estimate's precision. Otherwise, the coefficients may be under or overestimated, in addition to the increased likelihood of not finding an association even when it really exists. One should thus be aware of it when interpreting results in which the coefficient is not statistically significant, which will probably not allow concluding lack of association. Increasing the number of 24hR in at least a portion of the study population has a positive impact on the estimated coefficient's precision.

Contributors

E. Verly-Jr planned the study, conducted the statistical analyses, and wrote the manuscript. R. Sichieri and V. T. Baltar wrote and revised the manuscript.

Acknowledgments

The authors wish to acknowledge the funding from the Carlos Chagas Rio de Janeiro State Research Foundation (Faperj; grant n. E-26/201.488/2014) and the Brazilian National Council for Scientific and Technological Development (CNPq; n. 481434/2013-5).

References

1. Willett WC. Nutritional epidemiology. New York: Oxford University Press; 2012.
2. Kipnis V, Subar A, Midthune D. Structure of dietary measurement error: results of the OPEN Biomarker Study. Am J Epidemiol 2003; 158:14-21.
3. Schatzkin A, Kipnis V. Could exposure assessment problems give us wrong answers to nutrition and cancer questions? J Natl Cancer Inst 2004; 96:1564-5.
4. Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? Cancer Epidemiol Biomarkers Prev 2005; 14:2826-8.
5. Arab L, Wesseling-Perry K, Jardack P, Henry J, Winter A. Eight self-administered 24-hour dietary recalls using the Internet are feasible in African Americans and whites: the energetic study. J Am Diet Assoc 2010; 110:857-64.
6 Freedman LS, Schatzkin A, Thiebaut ACM, Potischman N, Subar AF, Thompson FE, et al. Abandon neither the food frequency questionnaire nor the dietary fat-breast cancer hypothesis. Cancer Epidemiol Biomarkers Prev 2007; 16;1321-2.
7. 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-61S.
8. Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, et al. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. J Am Diet Assoc 2006; 106:1575-87.
9. Nusser SM, Carriquiry AL, Dodd KW, Fuller WA. A semiparametric transformation approach to estimating usual daily intake distributions. J Am Stat Assoc 1996; 91:1440-9.
10. Freedman LS, Schatskin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. J Natl Cancer Inst 2011; 103:1086-92.
11. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. 2nd Ed. Boca Raton: Chapman and Hall/CRC Press; 2006.
12. Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, et al. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. Biometrics 2009; 65:1003-10.
13. Carroll RJ, Midthune D, Subar AF, Shumkovich M, Freedman LS, Thompson FE, et al. Taking advantages of the strengths of 2 different dietary assessment instruments to improve estimates for nutritional epidemiology. Am J Epidemiol 2012; 175:340-7.
14. Verly-Jr E, Castro MA, Fisberg RM, Marchioni DM. Precision of usual food intake estimates according to the percentage of individuals with a second dietary measurement. J Acad Nutr Diet 2012; 112:1015-20.
15. Conway JM, Ingwersen LA, Vinyard BT, Moshfegh A. Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and non obese women. Am J Clin Nutr 2003; 77:1171-8.
16. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. Statist Med 1998; 17:1623-34.
17. Nissenshon M, Román-Viñas B, Sánchez-Villegas A, Piscopo S, Serra-Majem L. The effect of the mediterranean diet on hypertension: a systematic review and meta-analysis. J Nutr Educ Behav 2016; 48:42-53.e1.
18. Zong G, Gao A, Hu FB, Sun Q. Whole grain intake and mortality from all causes, cardiovascular disease, and cancer. A meta-analysis of prospective cohort studies. Circulation 2016; 133:2370-80.
19. Cohen J. Statistical power analysis for the behavioral sciences. 2nd Ed. Mahwah: Lawrence Erlbaum Associates; 1988.
20. Szklo M, Javier Nieto F. Epidemiology: beyond the basics. Gaithersburg: Aspen; 2000.

Resumo

O uso de um ou poucos recordatórios de 24 horas (R24h) como medida da exposição dietética em modelos para estimar o efeito do alimento sobre um desfecho leva à atenuação da medida de associação e redução do poder do teste. Isso ocorre em função da variação dia a dia no consumo. A medida de associação pode ser corrigida por meio de calibração, e requer pelo menos uma repetição do R24 horas em uma subamostra da população de estudo. No entanto, seu desempenho e precisão dos coeficientes corrigidos podem variar de acordo com as características do estudo em que é aplicada. O objetivo deste trabalho é avaliar o desempenho da correção em diferentes cenários de pesquisa em relação à estimativa da medida de associação e sua precisão. Foi simulada população (n = 1.000) com informação sobre consumo de um alimento referente a 200 dias, e um desfecho com uma associação definida com o consumo usual (média dos 200 dias de consumo). Os cenários avaliados foram: (a) 100%, 60%, 40% e 20% da amostra com 2 dias de consumo; (b) indivíduos com 2, 3, 4 e 5 R24h; e (c) população com 1.000, 600 e 300 indivíduos. Os coeficientes foram estimados para 300 combinações aleatórias de dias de consumo; coeficientes corrigidos médios foram semelhantes ao verdadeiro coeficiente. A precisão foi menor em todos os cenários: a probabilidade de encontrar uma associação significativa (sendo ela verdadeira) variou de 0,47 a 0,29 (100% a 20% com repetição, respectivamente); 0,47 a 0,78 (2 a 5 dias); e 0,47 a 0,15 (1.000 a 300 indivíduos).

Inquéritos sobre Dietas; Consumo de Alimentos; Análise de Regressão; Nutrição em Saúde Pública

Resumen

El uso de uno o varios recuerdos de 24 horas (24hR) para medir la exposición dietética en modelos para estimar el efecto del alimento en un resultado conduce a la atenuación de la medida de asociación y una disminución en la potencia de la prueba. Esto se debe a la variación diaria en la ingesta de alimentos. La medida de asociación puede ser corregida por medio de calibración y requiere al menos una repetición del 24hR en una submuestra de la población de estudio. Sin embargo, el rendimiento de la calibración y la precisión de los coeficientes corregidos pueden variar de acuerdo con las características del estudio al que se aplica. El objetivo de este estudio es evaluar el desempeño de la corrección en diferentes escenarios de investigación en relación con la medida estimada de asociación y su precisión. Se simuló una población (n = 1.000) con información sobre la ingesta de alimentos durante 200 días y un resultado con una asociación definida con la ingesta habitual (media para los 200 días de ingesta). Los escenarios evaluados fueron: (a) 100%, 60%, 40% y 20% de la muestra con 2 días de ingesta; (b) individuos con 2, 3, 4 y 5 24hR; y (c) poblaciones con 1,000, 600 y 300 individuos. Los coeficientes se estimaron para 300 combinaciones aleatorias de días de ingesta; la media de los coeficientes corregidos fueron similares al verdadero coeficiente. La precisión fue menor en todos los escenarios: la probabilidad de encontrar una asociación significativa (cuando verdadera) varió de 0,47 a 0,29 (100% a 20% con repetición, respectivamente); 0,47 a 0,78 (2 a 5 días); y 0,47 a 0,15 (1.000 a 300 individuos).

Encuestas sobre Dietas; Consumo de Alimentos; Análisis de Regresión; Nutrición en Salud Pública

Submitted on 07/Oct/2016
Final version resubmitted on 05/Mar/2017
Approved on 22/Mar/2017