

Research

Open Access

## The use of income information of census enumeration area as a proxy for the household income in a household survey

Fabio S Gomes\*<sup>1,2</sup>, Mauricio TL Vasconcellos<sup>1</sup> and Luiz A Anjos<sup>3,4</sup>

Address: <sup>1</sup>Escola Nacional de Ciências Estatísticas, Fundação Instituto Brasileiro de Geografia e Estatística, Rua André Cavalcanti 106, Sala 403, Bairro de Fátima, 20231-050 - Rio de Janeiro, RJ, Brazil, <sup>2</sup>Área de Alimentação, Nutrição e Câncer, Coordenação de Prevenção e Vigilância, Instituto Nacional de Câncer, Rua dos Inválidos 212, 4° andar, Centro, 20231-048 - Rio de Janeiro, RJ, Brazil, <sup>3</sup>Laboratório de Avaliação Nutricional e Funcional, Departamento de Nutrição Social, Universidade Federal Fluminense, Rua Mario Santos Braga 30, Sala 415, Campus do Valonguinho, 24020-140 - Niterói, RJ, Brazil and <sup>4</sup>Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, Rua Leopoldo Bulhões 1480, 21041-210 - Rio de Janeiro, RJ, Brazil

Email: Fabio S Gomes\* - [fabiodasilvagomes@gmail.com](mailto:fabiodasilvagomes@gmail.com); Mauricio TL Vasconcellos - [mautlv@gmail.com](mailto:mautlv@gmail.com); Luiz A Anjos - [lanjos@gmail.com](mailto:lanjos@gmail.com)

\* Corresponding author

Published: 22 September 2009

Received: 12 December 2008

*Population Health Metrics* 2009, **7**:14 doi:10.1186/1478-7954-7-14

Accepted: 22 September 2009

This article is available from: <http://www.pophealthmetrics.com/content/7/1/14>

© 2009 Gomes et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Some of the Census Enumeration Areas' (CEA) information may help planning the sample of population studies but it can also be used for some analyses that require information that is more difficult to obtain at the individual or household level, such as income. This paper verifies if the income information of CEA can be used as a proxy for household income in a household survey.

**Methods:** A population-based survey conducted from January to December 2003 obtained data from a probabilistic sample of 1,734 households of Niterói, Rio de Janeiro, Brazil. Uniform semi-association models were adjusted in order to obtain information about the agreement/disagreement structure of data. The distribution of nutritional status categories of the population of Niterói according to income quintiles was performed using both CEA- and household-level income measures and then compared using Wald statistics for homogeneity. Body mass index was calculated using body mass and stature data measured in the households and then used to define nutritional status categories according to the World Health Organization. All estimates and statistics were calculated accounting for the structural information of the sample design and a significance level lower than 5% was adopted.

**Results:** The classification of households in the quintiles of household income was associated with the classification of these households in the quintiles of CEA income. The distribution of the nutritional status categories in all income quintiles did not differ significantly according to the source of income information (household or CEA) used in the definition of quintiles.

**Conclusion:** The structure of agreement/disagreement between quintiles of the household's monthly per capita income and quintiles of the head-of-household's mean nominal monthly income of the CEA, as well as the results produced by these measures when they were associated with the nutritional status of the population, showed that the CEA's income information can be used when income information at the individual or household levels is not available.

## Background

The place of health on the international agenda for development has been broadened [1] and health inequalities between and within countries have become a topic of great interest [2-5]. The concept of health inequalities includes the presence of unfair, avoidable, or remediable health differences among populations or specific groups defined according to social, economic, demographic or geographic criteria [6]. It implies a failure in avoiding or overcoming these differences that overlooks basic human rights [7].

For these reasons, it is common that population surveys collect socioeconomic information when the purpose is either exploratory or descriptive (and this information becomes the main focus) and socioeconomic information is associated with outcomes or other variables of interest.

Income and education are the most used variables to characterize and/or discriminate among socioeconomic groups. However, the collection of this information, particularly income, is sometimes difficult and can be influenced by other factors in population-based studies. These interferences may result in either total failure to obtain it or misreporting (under- or overestimation) [8].

In Brazil, the Census Enumeration Areas (CEA) are used to assess the data of the Brazilian Demographic Census but they are also used as conglomerates of households for other population-based surveys. They are defined as contiguous groups of approximately 300 households respecting administrative and political boundaries and identified by stable and easy location reference points [9]. Some of the CEA's information may help in planning the sample of such studies but it can also be used for some analyses that require information that is difficult to obtain at the individual or household level, such as income.

Although the use of this kind of information would be especially useful in developing countries, the few available studies in this area found in the literature were conducted exclusively in high-income countries in North America or Europe and in Australia [10-19]. There are remarkable differences in the methods of these studies, such as the definition of area levels, the independent and outcome variables adopted, and the statistical analysis, all of which hinder detailed comparisons. Most studies are interested in substituting individual-level [10,12,13,15-17,19], and more rarely household-level [14], information by the area level available in the census. Variables used to describe socioeconomic status vary from self-reported income data to socioeconomic scales. There is a variety of health outcomes in the analyses that make it difficult to generalize [10]. Furthermore, the studies analyze the data using different procedures such as factor analysis,

log-linear models, and estimation of correlation, agreement and reliability indexes such as intra-class correlation, Cohen's kappa or Kendall's coefficient of concordance. As a result, it is difficult to explain or predict the situations for which the different levels of socioeconomic information (e.g. CEA, household, individual) would produce similar results.

Additionally, no study has empirically compared the trade-offs in terms of cost savings, potential bias or loss of accuracy due to the use of area-level instead of individual- or household-level information. It has been suggested that the census-aggregated information is complementary because it may have a different construct meaning, depending on how it is defined in association with the health outcomes [10]. In middle- and low-income countries, savings would probably surpass bias and accuracy costs.

The gap between the year of the census (every 10 years in Brazil) and the year of a given survey may play a crucial role in the socioeconomic characteristics of the population. Additionally, the fact that some countries' economic growth may be stationary or there is very discrete social mobility may facilitate the comparisons because there may not be expressive changes in family or individuals' income or socioeconomic status between the year of the census and the survey. On the other hand, if the country's economic growth is reflected in individual and family income, one may not be able to use the census information.

The purpose of the present study was to assess the validity of household income data from CEA to represent household income obtained in a household survey. In practical terms, it sought to verify if the CEA income information could be used as a proxy for household income in a household survey conducted to assess the nutritional status of the population of Niterói, a city in the state of Rio de Janeiro, Brazil.

## Methods

The Nutrition, Physical Activity and Health Survey (*Pesquisa de Nutrição, Atividade Física e Saúde - PNAFS*) was the first household survey conducted to assess the nutritional status and health conditions of adolescents and adults living in Niterói, Rio de Janeiro, Brazil. Data collection was carried out between January and December 2003. Niterói is located in the metropolitan region of Rio de Janeiro that had 459,451 inhabitants in 2000, according to the last Brazilian census [9].

To guarantee the representativeness of the population of Niterói, a probabilistic sample of households was designed. The households were selected from the 2000

population CEA listing (there are 696 CEA with an average of 216 households per CEA in Niterói) [9] in two stages (CEA and household).

In the first stage, 110 CEA were selected, systematically, with probabilities proportional to the number of permanent private households. Prior to selection, the CEA were ordered from lowest to highest according to the head-of-household's mean nominal monthly income, thus implicitly stratifying the CEA by mean income and ensuring the selection of CEA from all income levels.

In the second stage, 16 households were selected in each CEA with equal probability, using an inverse sampling procedure [20], analogous to that applied in the World Health Survey in Brazil [21], leading to a sample size of 1,734 households (3,619 subjects) after the refusal of 26 households to participate in the study.

The sample weights were calculated as the product of inverse selection probabilities in each stage, using the estimator proposed by Haldane [20] adapted to be used in household surveys [21]. To reduce selection biases, common in household surveys, the natural sampling weights were calibrated to provide estimates that coincide with known population totals [22].

The calibration post-strata were defined using the variables age and sex. The combination of the two categories of sex (male and female) and age -- categorized as seven brackets: 10-19.9; 20-29.9; 30-39.9; 40-49.9; 50-59.9; 60-69.9; and 70 years or more -- resulted in 14 post-strata (2 sexes × 7 age brackets). For the calibration of sampling weights, the household natural weight ( $W_{ij}$ ) was multiplied by a calibration factor ( $g_{ij}$ ), providing the household calibrated weight  $W_{ij}^c = W_{ij} \times g_{ij} \forall ij \in d$ , where  $i$  represents the index of the selected CEA,  $j$  the index of the selected household and  $d$  the 14 post-strata domains, as indicated above. The Generalized Regression Estimator proposed by Deville & Särndal [23] was adopted to estimate the calibration factor  $g_{ij}$  as

$$g_{ij} = 1 + q_{ij} \left( \sum_{ij \in s} w_{ij} x_{ij} x_{ij}^t \right)^{-1} (t_x - \hat{t}_x)^t x_{ij},$$

where  $q_{ij}$  is a constant usually defined as 1 [22],  $x_{ij}$  represents the vector of auxiliary variables (i.e., sex and age),  $t_x$  denotes the vector of known population totals, and  $\hat{t}_x$  the vector with the estimates of the auxiliary variables calculated using natural sample weights.

Despite its extensive use, the Cohen's Kappa index ( $\kappa$ ) does not provide information about the agreement/dis-

agreement structure of data and it cannot be used to analyze ordinal scale categories [24], such as education or income strata. For this reason, the adjustment of uniform semi-association models was performed in the present analysis [25]. It is a generalized linear model of the Poisson family with *log* link function that considers the ordination structure of the variable's categories. Three components of the structure of agreement and disagreement compose this model: (1) the agreement at random; (2) the agreement due to the association between classifications; and (3) the agreement after eliminating the effects of the agreement at random and the association between variables [25,26]. Beyond the combination of the effects of agreement and disagreement, the semi-association model also considers the variations by categories, in the main diagonal, different from other models that assume the agreement is the same for each cell in the main diagonal [25,26].

Besides the adjustment of the model that assesses the agreement/disagreement between the classification of income categories defined according to the CEA or the field-obtained information on household income, two other models were adjusted: (1) the information on household income from male-headed and (2) female-headed households. This was motivated by the hypothesis that when the woman is the head of the household, she may not know exactly her spouse's income and vice versa. Therefore, household income might be estimated with different errors if the head of the household knows or does not know the spouse's income.

The adjusted model agreement grades were estimated for each cell in terms of *odds ratio* (OR), using the measure  $\tau_{ij}$  (where  $i$  indicates the line and  $j$  the column of the cell) as proposed by Darroch & McCloud [27].

In addition to the adjustment of the model, Cohen's weighted Kappa [28], Kendall's coefficient of concordance [29], Krippendorff's alpha reliability coefficient [30] and Spearman's correlation  $\rho$  [31] were also estimated in order to check the robustness of the model and to allow comparisons to other studies.

To illustrate the comparison between the two income information applied to an epidemiologic study, the distribution of nutritional status categories of the population of Niterói ( $\geq 10$  years of age) according to income quintiles (CEA and household) was performed.

To test the hypothesis that the distribution of the population by nutritional status categories according to the household income quintiles is equal to the distribution according to the quintiles constructed with the income of CEA the Wald statistic for homogeneity based in the sampling design was used [32].

**Table 1: Uniform semi-association model**

Parameter estimator	Parameter estimate	SE	p-value
$\hat{\lambda}$ (intercept)	9.23	0.010	< 0.001
$\hat{\lambda}_1^{household}$ (2 <sup>nd</sup> quintile of household income <sup>†</sup> )	-0.45	0.013	< 0.001
$\hat{\lambda}_2^{household}$ (3 <sup>rd</sup> quintile of household income <sup>†</sup> )	-0.80	0.013	< 0.001
$\hat{\lambda}_3^{household}$ (4 <sup>th</sup> quintile of household income <sup>†</sup> )	-1.51	0.016	< 0.001
$\hat{\lambda}_4^{household}$ (5 <sup>th</sup> quintile of household income <sup>†</sup> )	-2.24	0.021	< 0.001
$\hat{\lambda}_1^{CEA}$ (2 <sup>nd</sup> quintile of CEA's income)	-0.57	0.012	< 0.001
$\hat{\lambda}_2^{CEA}$ (3 <sup>rd</sup> quintile of CEA's income)	-1.22	0.013	< 0.001
$\hat{\lambda}_3^{CEA}$ (4 <sup>th</sup> quintile of CEA's income)	-1.91	0.018	< 0.001
$\hat{\lambda}_4^{CEA}$ (5 <sup>th</sup> quintile of CEA's income)	-3.22	0.023	< 0.001
$\beta$	0.31	0.002	< 0.001
$\delta_{i=1}$	0.38	0.026	< 0.001
$\delta_{i=2}$	0.51	0.024	< 0.001
$\delta_{i=3}$	0.10	0.024	< 0.001
$\delta_{i=4}$	0.81	0.022	< 0.001
$\delta_{i=5}$	0.39	0.017	< 0.001

<sup>†</sup>Quintiles of household income defined as the household's monthly per capita income.

SE = Standard Error

CEA = Census Enumeration Area (quintiles of CEA's income defined as the head-of-household's mean nominal monthly income of the CEA).

Body mass and stature data were collected in the households and used to calculate the body mass index (BMI = body mass in kilograms divided by stature in squared meters) as described elsewhere [33]. BMI for age and sex was used to define the nutritional status of adolescents (10-20 years of age) using the cut-off points presently recommended by the World Health Organization (WHO): low-BMI-for-age/thinness (< -2 Standard Deviations), overweight ( $\geq 1$  Standard Deviation) and obesity ( $\geq 2$  Standard Deviations) [34]. For adults ( $\geq 20$  years of age), the BMI cut-off points of < 18.5 kg/m<sup>2</sup>,  $\geq 25$  kg/m<sup>2</sup> and  $\geq 30$  kg/m<sup>2</sup> were used to define the categories of low-BMI/underweight, overweight and obesity, respectively [35].

The Institutional Review Board of the Sergio Arouca National School of Public Health of the Oswaldo Cruz Foundation approved all research procedures.

All estimates and statistics were calculated using the calibrated weights based on the structural information of the sample design, and a significance level lower than 5% was adopted. The analyses were conducted in R language and environment, version 2.6.1 [36].

### Results and Discussion

The parameters  $\hat{\lambda}_{1,2,3,4}^{household,CEA}$  represent the effect of a given quintile defined according to the household- or CEA-level income information on the log of the expected frequency

of being categorized in the first quintile by both household- and CEA-level income information ( $\hat{\lambda}$ ), which was set as reference to conduct the comparison. As the quintiles defined according to household- or CEA-level information get more distant from the reference quintile (first), the expected frequencies of being categorized in the first quintile according to one of the two levels of information and at the same time of being categorized in the second, then in the third, then in the fourth and then in the fifth quintile according to the other level of information, significantly decreases consecutively. These parameters account for the share of the agreement that is due to random (Tables 1 and 2).

The parameters  $\beta$  and  $\delta_i (i = j, i = 1 \dots, i)$ , estimated by  $\hat{\beta}$  and  $\hat{\delta}_i$  measure the association and agreement, respectively, between the measures of ordinal classification of income. The estimates of these parameters are statistically different from zero ( $p < 0.001$ ) (Table 1), which indicates that the assessments of income quintiles made by means of CEA and household information are not different and that the classification of households in the quintiles of household income tend to be associated with the classification of these households in the quintiles of CEA income. The agreement between measures ranges from 0

**Table 2: Uniform semi-association models adjusted using only household income information of households in which the head-of-household was male or female**

Parameter estimator	Sex of the head-of-household					
	Male			Female		
	Parameter estimate	SE	p-value	Parameter estimate	SE	p-value
$\hat{\lambda}$ (intercept)	8.45	0.015	< 0.001	8.62	0.013	< 0.001
$\hat{\lambda}_1^{household}$ (2 <sup>nd</sup> quintile of household income <sup>†</sup> )	-0.33	0.018	< 0.001	-0.57	0.018	< 0.001
$\hat{\lambda}_2^{household}$ (3 <sup>rd</sup> quintile of household income <sup>†</sup> )	-0.73	0.018	< 0.001	-0.87	0.020	< 0.001
$\hat{\lambda}_3^{household}$ (4 <sup>th</sup> quintile of household income <sup>†</sup> )	-1.44	0.022	< 0.001	-1.62	0.025	< 0.001
$\hat{\lambda}_4^{household}$ (5 <sup>th</sup> quintile of household income <sup>†</sup> )	-2.11	0.028	< 0.001	-2.54	0.033	< 0.001
$\hat{\lambda}_1^{CEA}$ (2 <sup>nd</sup> quintile of CEA's income)	-0.70	0.015	< 0.001	-0.42	0.018	< 0.001
$\hat{\lambda}_2^{CEA}$ (3 <sup>rd</sup> quintile of CEA's income)	-1.48	0.018	< 0.001	-0.94	0.020	< 0.001
$\hat{\lambda}_3^{CEA}$ (4 <sup>th</sup> quintile of CEA's income)	-2.38	0.025	< 0.001	-1.45	0.025	< 0.001
$\hat{\lambda}_4^{CEA}$ (5 <sup>th</sup> quintile of CEA's income)	-3.90	0.033	< 0.001	-2.56	0.033	< 0.001
$\beta$	0.36	0.003	< 0.001	0.27	0.003	< 0.001
$\delta_{i=1}$	0.40	0.036	< 0.001	0.34	0.037	< 0.001
$\delta_{i=2}$	0.94	0.032	< 0.001	-0.11	0.037	< 0.01
$\delta_{i=3}$	0.09	0.034	< 0.05	0.10	0.034	< 0.01
$\delta_{i=4}$	0.91	0.030	< 0.001	0.78	0.033	< 0.001
$\delta_{i=5}$	0.60	0.024	< 0.001	0.14	0.025	< 0.001

<sup>†</sup>Quintiles of household income defined as the household's monthly per capita income.

SE = Standard Error

CEA = Census Enumeration Area (quintiles of CEA's income defined as the head-of-household's mean nominal monthly income of the CEA).

(no agreement) to 1 (perfect agreement). All agreement estimates ( $\hat{\delta}_i$ ) are significantly different from zero, which means that there is enough statistical evidence to reject the hypothesis that there is no agreement between the pairs of quintiles defined by household- and CEA-level information (Tables 1 and 2).

Table 2 shows the estimates of the cited parameters for male-headed and female-headed households. The uniform semi-association models adjusted for the distinct sexes has resulted in the same conclusion suggesting that the sex of the head of the household does not influence the structure of agreement/disagreement of the information of CEA and household income.

Based on the model presented in Table 1, the agreement grades  $\hat{\tau}_{ij}$  were estimated for each  $ij$  cell and its respective 95% confidence interval (95% CI) (Table 3). Adopting the first cell of the diagonal ( $i = 1, j = 1$ ) as an example of interpretation, the value in the matrix's diagonal represents the OR of a household to be categorized in the 1<sup>st</sup>

quintile of income by the CEA income is two times greater ( $\hat{\tau}_{11} = 2.1$ ; 95% CI = 1.9-2.4) when the household is located in the 1<sup>st</sup> quintile of household income. To illustrate the interpretation of values that are out of the diagonal, the cell located in line 1, column 5 was taken. In this case, the OR of a household to be categorized in the 1<sup>st</sup> quintile of household income rather than in the 5<sup>th</sup> quintile of household income is approximately 308 times greater ( $\hat{\tau}_{15} = \hat{\tau}_{51} = 307.9$ ; 95% CI = 263.6-359.7), considering that the household, according to the CEA income, belongs to the 1<sup>st</sup> quintile rather than the 5<sup>th</sup> quintile.

The values of  $\hat{\tau}_{ij}$  indicates the OR of assessment measures to be concordant rather than discordant. Observing the values in the first line (fixing line 1 and varying columns) or in the fifth column (fixing column 5 and varying lines) of Table 3, it is possible to note that the agreement increases as the quintiles get more distant from the others, according to the interpretation above (Table 3).

According to the agreement classifications more widely used [37,38], the estimated indexes indicate fair to moderate levels of agreement and are statistically significant:  $Kappa_w = 0.49$  ( $p < 0.001$ ); Kendall's coefficient of concordance = 0.41 ( $p < 0.001$ ); Krippendorff's alpha = 0.48; Spearman's  $\rho = 0.49$  ( $p < 0.001$ ).

Figure 1 presents the proportional distribution of nutritional status categories for each income quintile defined according to the household's monthly per capita income and the head-of-household's mean nominal monthly income of the CEA. The homogeneity between the pairs of proportions' vectors of each income quintile indicates that the distribution of the nutritional status categories in all income quintiles does not differ significantly according to the source of income information (household or CEA) used in the definition of quintiles (Figure 1). In other words, from a multivariate perspective, there is not enough statistical evidence to reject the hypothesis that the proportional distribution of nutritional status categories within each quintile would be different according to the source of income information used to categorize the quintiles.

The results of studies that investigate the use of area-level socioeconomic information as proxy of household or individual information are still controversial regarding the agreement of income measures as well as the results produced by each measure when related with an outcome. This is expected because the analysed outcomes, methods

employed in the definition of socioeconomic strata, and the partitioning criteria used to define the territories vary between studies [12,16,39].

On one hand, the literature indicates that the information of both levels can be used without jeopardizing the analyses on health inequities because they produce similar results [15,16,18]. On the other hand, there is also evidence that the use of area-level information could result in substantial errors in the classification of socioeconomic conditions and, therefore, could not predict certain health outcomes as well as individual-level information [10-14,17,19].

In the present study, the structure of agreement/disagreement between quintiles of household monthly per capita income and quintiles of the head-of-household's mean nominal monthly income of the CEA, as well as the results produced by these measures when they were associated with the nutritional status of the population of Niterói, showed that the CEA's income information can be used when income information at the individual or household level are not available.

The hypothesis that the sex of the head of the household would not influence the structure of agreement/disagreement of income categories could not be rejected. Other factors, such as race [10,39], that could influence this structure were not analyzed in the present study. Another limitation of this analysis consists of the definition of par-

**Table 3: Symmetric matrix with the agreement grades estimates ( $\hat{\tau}_{ij}$ ) for each cell, with its respective 95% confidence intervals (95% CI).<sup>§‡</sup>**

Income quintiles	CEA				
	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>
<i>Household</i> <sup>†</sup>					
1 <sup>st</sup>	2.1 (1.9-2.4)	3.3 (3.0-3.7)	5.6 (5.0-6.3)	53.7 (47.0-61.3)	307.9 (263.6-359.7)
2 <sup>nd</sup>		2.8 (2.5-3.0)	2.5 (2.3-2.8)	13.0 (11.6-14.4)	40.0 (35.4-45.1)
3 <sup>rd</sup>			1.2 (1.1-1.4)	3.4 (3.1-3.7)	5.7 (5.1-6.2)
4 <sup>th</sup>				5.1 (4.7-5.5)	4.5 (4.2-4.9)
5 <sup>th</sup>					2.2 (2.0-2.3)

All  $\hat{\tau}_{ij}$  are significantly different from 1.

<sup>§</sup>In the diagonal,  $\hat{\tau}_{ij}$  represents the Odds Ratio of a household to be categorized in the same quintile by both household and CEA income.

<sup>‡</sup> Out of the diagonal,  $\hat{\tau}_{ij}$  represents the Odds Ratio of a household to be categorized in the  $i^{th}$  quintile of household income rather than in a different ( $j^{th}$ ) quintile of household income, considering that the household, according to the CEA income, belongs to the  $j^{th}$  quintile rather than the  $i^{th}$  quintile.

<sup>†</sup>Quintiles of household income defined as the household's monthly per capita income.

CEA = Census Enumeration Area (quintiles of CEA's income defined as the head-of-household's mean nominal monthly income of the CEA).

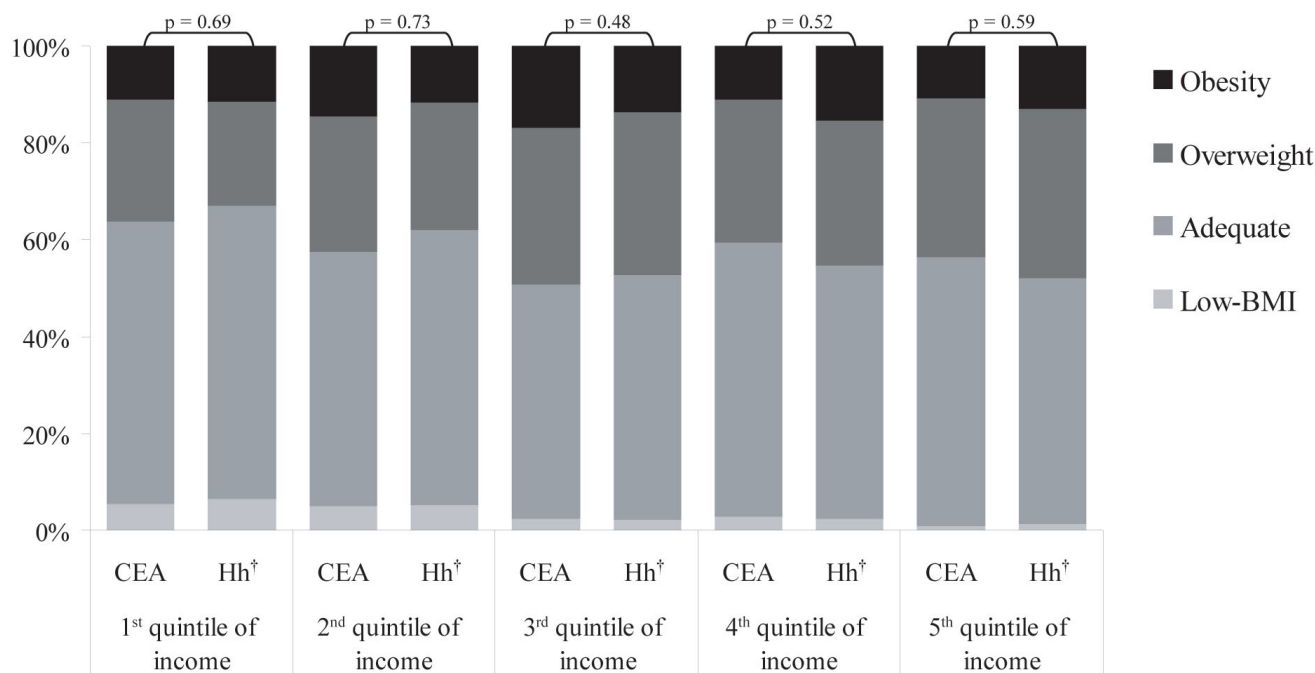
tioning of income strata in quintiles. This procedure was adopted based on an applied criterion, with an analytic purpose, since partitions with large intervals are generally used only with the purpose of planning sample and/or study design. However, it is important to register that partitions in larger intervals could have led to different conclusions because the distance between categories would be reduced, which could result in greater classification errors. On the other hand, smaller partitions -- in thirds, for example -- would increase the distance between categories, diminishing the chances of occurrence of misclassifications.

It is also important to pay attention to the time between when the information were assessed. This is particularly important in countries undergoing fast economic growth or greater socioeconomic mobility. The survey used in the present analysis was conducted only three years after the 2000 Brazilian census [9]. During this three-year period, the Brazilian economy was stable due to inexpressive and unsustainable economic growth, and socioeconomic mobility was also compromised by declining gross domestic product, wage-share and continuous reduction of the formal employment sector.

Additionally, it is also important to note that the aggregated census information comes from individually collected information, which may raise the question whether the individual information collected in the census is reliable. The census information cannot be regarded as gold standard but it is expected to constitute more robust information than that collected in surveys because there are many more quality control mechanisms, proportionally fewer missing values, higher trust in the interviewer as an employee of a known institution (the Census Bureau), and no variance due to sample design.

Another issue that could be raised when dealing with aggregated data is that the income distribution within a given aggregated level (e.g., CEA, city) may vary according to the distance from a predetermined center. However, the adjusted models have not taken into account the modifiable areal unit problem and ecological fallacies due to aggregation [40-42], a limitation of the present study due to the absence of information on the distance between households and a predetermined center.

Furthermore, the inference and conclusions of this study may not apply to different variables of interest, countries,



**Figure 1**  
**Proportional distribution of the Niterói population (≥ 10 years of age), according to the nutritional status, by quintiles of household's monthly per capita income and quintiles of the head-of-household's mean nominal monthly income of the CEA.** †Quintiles of household (Hh) income defined as the household's monthly per capita income. CEA = Census Enumeration Area (quintiles of CEA's income defined as the head-of-household's mean nominal monthly income of the CEA). p = p-value obtained from the Wald statistics for homogeneity between pairs of proportion vectors according to the nutritional status by income quintiles (α = 0.05).

sizes and boundaries of enumeration areas, and possibly survey designs. Therefore, comparisons by other studies should be carefully made, taking this limitation into account.

## Conclusion

It is remarkable that until this paper, the few studies on this theme had been solely derived from high-income countries (United States of America [10-12], Canada [13-16], Italy [17], Spain [18] and Australia [19]). For this reason, it would be important to encourage other analyses using different levels of aggregation and territories in low- and middle-income countries such as Brazil, which would make intra- and international comparison possible, contributing to the collection of evidence about the use of socioeconomic information aggregated by CEA in the absence of individual information.

This is perhaps the first study conducted in a developing country that compares the use of area- versus household-level income measures in association with a health outcome (nutritional status). The study indicates that CEA's income information may be used as a proxy for household income in the absence of individual- or household-level information. The sex of the source of household income information did not influence the structure of agreement/disagreement of income categories. Additionally, the association between income quintiles and nutritional status is similar whether CEA- or household-level income measures were used.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

LAA and MTLV planned the research. MTLV designed the sample. FSG and MTLV calculated the natural and calibrated sampling weights and performed the statistical analysis. FSG wrote the first draft of the paper, which was revised and approved by the other authors.

## Acknowledgements

The Nutrition, Physical Activity, and Health Survey was partially funded by the Brazilian National Research Council (Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq; grants nos. 471172/2001-4 and 475122/2003-8) and by the Oswaldo Cruz Foundation (PAPES III - Program to Support Strategic Projects in Health, no. 250.139). LAA received a research productivity grant from CNPq (grant no. 301076/89-8 and 311801/06-4). MTLV received a research productivity grant from CNPq (grant no. 302992/2003-0).

## References

- World Health Organization (WHO), Secretariat of the Commission on Social Determinants of Health: **Action on the social determinants of health: learning from previous experiences**. Geneva 2005.
- WHO: **The World Health Report 2003. Shaping the future**. Geneva 2003.
- Evans T, Whitehead M, Diderichsen F, Bhuiya A, Wirth M, Eds: *Challenging inequities in health: from ethics to action* New York: Oxford University Press; 2001.
- Leon DA, Walt G, Eds: *Poverty, inequality and health: an international perspective* New York: Oxford University Press; 2001.
- Kim JY, Millen JV, Irwin A, Gershman J, Eds: *Dying for growth: global inequality and the health of the poor* Monroe: Common Courage Press; 2000.
- WHO, Executive Board EB115/35: **Note by the Secretariat**. In *115th Session: 25 November 2004*; Geneva the Commission on Social Determinants of Health: WHO; 2004:1-3.
- Graham H, Kelly MP: *Health inequalities: concepts, frameworks and policy 2004* [[http://www.nice.org.uk/niceMedia/pdf/health\\_inequalities\\_policy\\_graham.pdf](http://www.nice.org.uk/niceMedia/pdf/health_inequalities_policy_graham.pdf)]. NHS Health Development Agency
- Barros RP, Cury S, Ulyseia G: *A desigualdade de renda no Brasil encontra-se subestimada? Uma análise comparativa com base na PNAD, na PO e nas Contas Nacionais* Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada; 2007.
- Instituto Brasileiro de Geografia e Estatística: *Censo Demográfico 2000: agregado por setores censitários dos resultados do universo* Rio de Janeiro; 2003.
- Diez Roux AV, Kiefe CI, Jacobs DR Jr, Haan M, Jackson SA, Nieto FJ, Paton CC, Schulz R: **Area characteristics and individual-level socioeconomic position indicators in three population-based epidemiologic studies**. *Ann Epidemiol* 2001, **11**:395-405.
- Diez Roux AV, Merkin SS, Hannan P, Jacobs DR, Kiefe CI: **Area characteristics, individual-level socioeconomic indicators, and smoking in young adults**. *Am J Epidemiol* 2003, **157**:315-326.
- Geronimus AT, Bound J: **Use of census-based aggregate variables to proxy for socioeconomic group: evidence from national samples**. *Am J Epidemiol* 1998, **148**:475-486.
- Demissie K, Hanley JA, Menzies D, Joseph L, Ernst P: **Agreement in measuring socio-economic status: area-based versus individual measures**. *Chronic Dis Can* 2000, **21**:1-7.
- Hanley GE, Morgan S: **On the validity of area-based income measures to proxy household income**. *BMC Health Services Research* 2008, **8**:1472-6963.
- Janssen I, Boyce WF, Simpson K, Pickett W: **Influence of individual- and area-level measures of socioeconomic status on obesity, unhealthy eating, and physical inactivity in Canadian adolescents**. *Am J Clin Nutr* 2006, **83**:139-145.
- Southern DA, McLaren L, Hawe P, Knudtson ML, Ghali WA: **Individual-level and neighborhood-level income measures: agreement and association with outcomes in a cardiac disease cohort**. *Med Care* 2005, **43**:1116-1122.
- Cesaroni G, Farchi S, Davoli M, Forastiere F, Perucci CA: **Individual and area-based indicators of socioeconomic status and childhood asthma**. *Eur Respir J* 2003, **22**:619-624.
- Dominguez-Berjón F, Borrell C, Rodríguez-Sanz M, Pastor V: **The usefulness of area-based socioeconomic measures to monitor social inequalities in health in Southern Europe**. *Eur J Public Health* 2006, **16**:54-61.
- Walker AE, Becker NG: **Health inequalities across socio-economic groups: comparing geographic-area-based and individual-based indicators**. *Public Health* 2005, **119**:1097-1104.
- Haldane JBS: **On the method of estimating frequencies**. *Biometrika* 1945, **33**:222-225.
- Vasconcellos MT, Silva PL, Szwarcwald CL: **Sampling design for the World Health Survey in Brazil**. *Cad Saude Publica* 2005, **21**:89-99.
- Silva PLN: *Calibration estimation: when and why, how much and how* Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2004.
- Deville JC, Särndal CE: **Calibration estimators in survey sampling**. *J Am Stat Assoc* 1992, **87**:376-382.
- Cohen J: **A coefficient of agreement for nominal scales**. *Educ Psychol Meas* 1960, **20**:37-46.
- Goodman LA: **Simple models for the analysis of association in cross-classifications having ordered categories**. *J Am Stat Assoc* 1979, **74**:537-552.
- Silva EF, Pereira MG: **Rating of the structures of agreement and disagreement in reliability studies**. *Rev Saude Publica* 1998, **32**:383-393.
- Darroch J, McCloud Pl: **Category of distinguishability and observer agreement**. *Aust J Stat* 1986, **28**:371-388.



28. Cohen J: **Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit.** *Psychol Bull* 1968, **70**:213-220.
29. Kendall MG: *Rank correlation methods* 4th edition. London: Charles Griffin & Company; 1970.
30. Krippendorff K: *Content analysis: an introduction to its methodology* Beverly Hills: Sage; 1980.
31. Spearman C: **The proof and measurement of association between two things.** *Amer J Psychol* 1904, **15**:72-101.
32. Pessoa DGC, Silva PLN, Eds: *Análise de dados amostrais complexos* São Paulo: Associação Brasileira de Estatística; 1998.
33. Bossan FM, Anjos LA, Vasconcellos MTL, Wahrlich V: **Nutritional status of the adult population in Niterói, Rio de Janeiro, Brazil: the Nutrition, Physical Activity, and Health Survey.** *Cad Saude Publica* 2007, **23**:1867-1876.
34. De Onis M, Onyango AVW, Borghi E, Siyam A, Nishida C, Siekmann J: **Development of a WHO growth reference for school-aged children and adolescents.** *Bull World Health Organ* 2007, **85**:660-667.
35. WHO: **Physical status: The Use and Interpretation of Anthropometry. Report of a WHO Expert Committee. Technical Report Series 854.** Geneva 1995.
36. R Development Core Team: *R: A language and environment for statistical computing* Vienna: R Foundation for Statistical Computing; 2007.
37. Shrout PE: **Measurement reliability and agreement in psychiatry.** *Stat Methods Med Res* 1998, **7**:301-317.
38. Landis JR, Koch GG: **The measurement of observer agreement for categorical data.** *Biometrics* 1977, **33**:159-174.
39. Braveman PA, Cubbin C, Egerter S, Chideya S, Marchi KS, Metzler M, Posner S: **Socioeconomic status in health research: one size does not fit all.** *JAMA* 2005, **294**:2879-2888.
40. Openshaw S: **Ecological fallacies and the analysis of areal census data.** *Environ Plan A* 1984, **16**:17-31.
41. Openshaw S: *The modifiable areal unit problem. Concepts and techniques in modern geography* Norwich: Geo Books; 1984.
42. Tagashira N, Okabe A: **The modifiable areal unit problem in a regression model whose independent variable is a distance from a predetermined point.** *Geogr Anal* 2002, **34**:1-20.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

