UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE ECONOMIA
PIMES - PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

SAMMARA CAVALCANTI SOARES

ENSAIOS SOBRE DESIGUALDADE EM SAÚDE AUTO AVALIADA NO BRASIL

Recife
2012

SAMMARA CAVALCANTI SOARES

ENSAIOS SOBRE DESIGUALDADE EM SAÚDE AUTO AVALIADA NO BRASIL

Dissertação apresentada ao Programa de Pós-Graduação em Economia – PIMES da Universidade Federal de Pernambuco para obtenção do título de Mestre em Economia sob a orientação da Profª Drª Tatiane Almeida de Menezes e co-orientação do Prof. Dr. Antonio Paez

Recife
2012

Catalogação na Fonte
Bibliotecária Ângela de Fátima Correia Simões, CRB4-773

S676e Soares, Sammara Cavalcanti
Ensaios sobre desigualdade em saúde auto avaliada no Brasil / Sammara Cavalcanti Soares. - Recife : O Autor, 2012.
87 folhas : il. 30 cm.

Orientador: Profª. Dra. Tatiane Ameida de Menezes e co-orientador Prof. Dr. Antonio Paez.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCSA. Economia, 2012.
Inclui bibliografia e anexos.

1. Índice de desigualdade em saúde. 2. Saúde auto avaliada. 3. Modelo probit ordinal. I. Menezes, Tatiana Almeida de (Orientador). II. Paez, Antonio(Co-orientador). III. Título.

331 CDD (22.ed.) UFPE (CSA 2012 – 147)

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE ECONOMIA
PIMES/PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE DISSERTAÇÃO DO MESTRADO ACADÊMICO EM ECONOMIA DE:

SAMMARA CAVALCANTI SOARES

A Comissão Examinadora composta pelos professores abaixo, sob a presidência do primeiro, considera o Candidato Sammara Cavalcanti Soares **APROVADO**.

Recife, 10/08/2012

Profa. Dra. Tatiane Almeida de Menezes
Orientadora

Prof. Dr. Raul da Mota Silveira Neto
Examinador Interno

Profa. Dra. Mônica Viegas Andrade
Examinador Externo

AGRADECIMENTOS

Gostaria de manifestar minha enorme gratidão primeiramente ao nosso Senhor Pai, por cada amanhecer, e junto com ele, a Sua bênção sobre nós;

À minha família amada, por ser minha base, minha força, meu equilíbrio;

À minha orientadora querida Tatiane Almeida de Menezes, por todos os ensinamentos, atenção e paciência. Considero-a acima de tudo uma verdadeira amiga, a qual tenho tanta admiração e que devo boa parte do que aprendi na minha caminhada acadêmica;

Ao meu co-orientador Antonio Paez, por ter me proporcionado ensinamentos e ideias fundamentais ao desenvolvimento do meu trabalho; pela oportunidade única de ter ido ao Canadá fazer parte do time da McMaster University; e de ter tido a atenção, acolhimento e amizade não só dele, mas também da sua linda família;

Ao Governo do Canadá, que através da bolsa ELAP, me concedeu uma experiência singular no seu país, e uma oportunidade de desenvolver não apenas minhas habilidades acadêmicas, mas também pessoais;

À imprescindível ajuda de Pedro Luís Silva e Marcus Vinícius Fernandes, da equipe do IBGE, que me forneceram informações essenciais sobre o banco de dados da PNAD;

A amigos e colegas que de alguma forma participaram e ajudaram em todo esse processo. Em especial, Monir Moniruzzaman, Kristina Cimaroli, Synthia Santana, Edivaldo Júnior, Rodrigo Arruda, Thiago Bananeiras, Cecília Lins, Luíza Campos.

Por fim, dedico esse trabalho ao meu namorado maravilhoso, Thiago Zilli, por toda a força, companhia, amor, amizade e confiança. Por ter sido a companhia mais essencial durante toda essa trajetória. Se eu consegui superar todas as dificuldades enfrentadas, e chegar até aqui, é a ele que direciono com todo amor, meu especial agradecimento.

RESUMO

O primeiro ensaio propõe uma nova abordagem para estimar desigualdades socioeconômicas na Saúde Auto Avaliada. O método baseia-se em uma aplicação alternativa do Coeficiente de Gini, preservando a natureza categórica da variável e evitando incorrer nas dificuldades e limitações de cardinalizar tal indicador de saúde para análises de desigualdade. O proposto "Index-D" aplica as probabilidades preditas de um Modelo Probit Ordinal, sob a equação de Gini formatada para funções densidade discretas, permitindo-nos estimar variações na Saúde Auto Avaliada segundo subjacentes variações socioeconômicas e demográficas. Usamos dados da Pesquisa Nacional por Amostra de Domicílio (PNAD) para os anos de 1998, 2003 e 2008, apenas referente à população feminina, a fim de ilustrar a aplicabilidade do método. Os resultados mostram que a desigualdade em Saúde Auto Avaliada no Brasil decresceu de 1998 a 2008 entre as mulheres, independentemente do perfil socioeconômico considerado. Ainda, os grupos femininos com melhores condições financeiras apresentaram índices de desigualdade menores, enquanto que os mais pobres obtiveram os maiores escores. Considerando a desigualdade entre as Regiões, sudeste apresentou os resultados mais favoráveis, enquanto o Norte e o Nordeste reportaram as mais altas desigualdades, independentemente do ano e do perfil socioeconômico considerado.

O segundo ensaio visa identificar o quanto da variância observada na Saúde Auto Avaliada (SAH) no Brasil é resultado do contexto onde as pessoas vivem. Dessa forma, através do *Random-Intercept Ordered Probity Model,* aplicamos uma amostra de municípios, retirada da Pesquisa Nacional por Amostra de Domicílios (PNAD), 2008, para representar as unidades do segundo nível, juntamente com as informações socioeconômicas dos indivíduos, a fim de controlar apropriadamente o efeito composição. Apesar de pequeno, o coeficiente de variação mostra a existência de variação sistemática na Saúde Auto Reportada entre os municípios urbanos do Brasil que persistiram mesmo após o controle do nível individual. As evidências sugerem que políticas de saúde no Brasil não devem investir apenas nas circunstâncias a nível individual, mas também sobre os ambientes sociais e físicos do coletivo, tais como segurança, espaços para lazer e infraestrutura urbana.

Palavras Chave: Índice de desigualdade em saúde; Saúde Auto Avaliada; Modelo Probit Ordinal; Modelos Multiníveis; Efeito Contextual; Municípios.

ABSTRACT

The first paper proposes a new approach to estimate socioeconomic inequalities in Self Assessed Health. The method relies on an alternative application of Gini Coefficient that preserves the categorical nature of the Self Assessed Health variable, and as a result, do not incur in the difficulties and limitations of cardinalizing this health indicator for inequality analysis purposes. The proposed “Index-D”, employs the predicted probabilities of an Ordinal Probit Model at a Gini equation fitted for discrete density function, in order to assess variations in Self-Assessed Health due to variations in the underlying socio-economic and demographic attributes. We use data from Brazilian National Household Surveys (PNAD) of 1998, 2003, and 2008 years, referring only to the women population, to illustrate its applicability. The results show that Self-Assessed Health inequality decreased in Brazil from 1998 to 2008 independently of the regarded socioeconomic profile. Also, the Within Index-D reported that, the better-off women groups hold lower health inequality; whereas the poorest population presented greater scores. Concerning the inequality across the regions, southeast obtained the most favorable results, while North and Northeast reported the highest health inequalities, regardless the year and the socioeconomic groups considered.

The objective of the second paper relies on identify in what extent the variability on Self Assessed Health (SAH) in Brazil is a result of the context where the people live. Thus, through a *Random-Intercept Ordered Probity Model*, we employ a sample of municipalities from 2008’s Brazilian National Household Survey (PNAD) to represent our secondary level units; along with individuals´ socioeconomic information, to properly control the potential compositional effect. Although small, the results show the existence of systematic SAH variance between the urban centers in Brazil that persisted even after individual-level control. The evidences suggest that health policy in Brazil should invest not only on circumstances at individual level, but also on the collective social- physical environments – such as security, spaces for leisure and urban facilities.

Key Words: Health Inequality Index; Self Assessed Health; Ordered Probit Model; Multilevel Models; Contextual Effect; Municipalities.

LISTA DE TABELAS

LISTA DE FIGURAS

SUMÁRIO

1. ASSESSING INEQUALITIES IN SELF-ASSESSED HEALTH BASED ON THE PROBABILITIES OF AN ORDINAL MODEL: A CASE STUDY OF INDIVIDUAL AND REGIONAL VARIATIONS IN BRAZIL

1.1 INTRODUCTION

Good health is a precondition for the development of functional capabilities, including education, employment, recreation, and in general, the conduct of activities that increase personal welfare and contribute to society (McCall et al. 2011). Not surprisingly, health is increasingly recognized as a fundamental aspect in the promotion of equitable and sustainable development for countries and regions, by international organizations (see the Millennium Development Goals of the United Nations[1]; the World Bank, 2007), and national governments alike (Brazilian National Commission on Social Determinants of Health, Final Report, 2008).

Health is a state of physical, mental, and social well-being that is influenced by factors in multiple dimensions: individual preferences, biological and environmental factors, and socio-economic and demographic conditions (e.g. Noronha 2002). Preferences are related to habits, behavioral patterns and choices of individuals, such as smoking, diet, and exercising. Biological and environmental factors are related to genetic constitution, exposure to risk, and illnesses associated with the life cycle. Other aspects related to socio-economic status (SES) and demographic factors, include income, occupational status, educational attainment, race and ethnic origin, marital status, and place of residence. Each of these dimensions is important from a public health perspective. Preferences can be related to efforts to curb habits that negatively impact health (e.g. smoking) or promote those that lead to favorable health outcomes (e.g. exercise). Accident prevention and the treatment of illness are also key aspects of public health. In the context of assessing the status and progress of development objectives, a thorough understanding of the effect of socio-economic and demographic factors is particularly valuable, in order to inform the design and targeting of policies aimed at promoting equitable public health.

Along these lines, there is consistent evidence throughout the world that the poor are afflicted by higher rates of illness and mortality than the relatively better-off (see Wagstaff, 2000 and World Bank, 1997). This is true between countries, with life expectancies in Japan

[1] http://web.undp.org/mdg/index.shtml

and France exceeding 80 years, while reaching only 68 and 45 in Bolivia and Afghanistan respectively. Gaps in health status between socio-economic and demographic groups can also be very pronounced within nations as well. In the case of Brazil, for instance, the infant mortality rate (IMR) despite has decreased over the last two decades, still there were 30.57 deaths per 1,000 live births by 2000 (Barufi et al, 2012). The imperative of alleviating inequalities in health as they relate to socio-economic status and demographic situation has prompted academic research and policy discussions in recent years. Several empirical (e.g. Kakwani 1994; Wagstaff and van Doorslaer 2003; Lindeboom et. al 2004) and theoretical (e.g. Gwatkin 2000; Adler and Ostrove,1999) works have focused on the matter of assessing the socio-economic determinants of inequalities in health outcomes. The results of these investigations tend to indicate a link between socio-demographic factors and health, for a majority of health indicators and regions investigated (Humphries and van Doorslaer, 2000).

Research on the SES and demographic determinants of health status has been conducted using a variety of objective indicators of health, such as mortality, morbidity, and functional limitations (see Wagstaff, 2000; Tubeuf and Perronnin, 2008; Yiengprugsawan et.al. 2010). Other research, in contrast, has relied on self-reported data, as for instance BMI (see Sanchez et. al., 1999) or on individual assessments of health condition, such as Self-Assessed Health (SAH) (see Doorslaer and Jones, 2002; Wagstaff 2001; Humphries and Doorslaer, 2000). The latter variable, recorded on an ordinal scale (i.e. “poor”, “fair”, “good”, “very good”, and “excellent”), provides a summary indicator of the respondent’s own perception of her general health condition. SAH has been shown to be a good predictor of various health outcomes, such as morbidity and mortality (see Idler and Benyamini, 1997; van Doorslaer and Gerdtham, 2003). This measure is convenient, easy to collect, and is in fact recorded in many contexts, on a sufficiently broad basis to allow for comparisons related to inequality.

On the other hand, SAH presents other challenges. For instance, for a given true health status, individuals with different demographic and socio-economic background are expected to use different reference points when asked to assess their health. This is the so-called reporting heterogeneity (Lindeboom and van Doorslaer 2004) or state-dependent reporting bias (Kerkhofs and Lindeboom, 1995). This reporting heterogeneity potentially implies that members of different socio-economic and demographic groups may perceive their own health differently even if the “true” health status is the same.

In order to use SAH in inequality analysis, a number of approaches have been proposed to transform SAH into a cardinal variable. These include the use of interval regression, the standard lognormal approach, and the use of the latent variable of an ordinal model (e.g. Waggstaff, 1994; van Doorslaer and Jones, 2002; Madden, 2008; Shmueli 2003). Despite their ability to generate a summary measure that is comparable across individuals, the appropriateness of enforcing cardinality on what essentially is a categorical variable has been questioned (e.g. Apouey, 2007). Therefore, the objective of this paper is to propose a novel approach to assess inequality in Self-Assessed Health. The approach relies on the use of an ordinal model, and the use of the probabilities obtained from such a model to generate a measure that can be compared to assess variations in Self-Assessed Health due to variations in the underlying socio-economic, demographic, and other relevant factors. As we will show, the approach has a number of attractive features. The measure that we propose does not change the original properties of the SAH, but rather employs those properties without imposing cardinality. The measure is bounded between zero and one, which facilitates interpretation. It is also easy to implement, and flexible to deploy for various profiles to allow several investigations of inequality. Since it is based on the estimated probabilities, it can be used to assess, in addition to inter-class inequality, the degree of intra-class inequality as well.

We demonstrate our approach by means of a case study of Self-Assessed Health in Brazil. Data obtained from the National Household Surveys of 1998, 2003, and 2008, allows us to assess variations in inequality in Self-Assessed Health by individual profile and region of the country for these years. The rest of the paper is organized as follows. In the following section we review the technical background and introduced our proposed approach. Next, we describe the context for the case study, and present and discuss the results of the analysis. We conclude with some final remarks about the applicability and potential of the approach.

1.2 METHOD

1.2.1 SELF-ASSESSED HEALTH AND APROACHES TO ASSESS INEQUALITY

Many variables of interest from a public health perspective are recorded using categories, as opposed to continuous variables. A prime example of this is Self-Assessed Health, a variable that is recorded in an ordinal scale, typically using a five-point scale ranging from "poor" to "excellent". It has been noted that the ordinal nature of this variable presents a challenge for the analysis of inequality, since most conventional measures of inequality, such as the use of concentration curves, calculation of Gini coefficient, and Atkinson measures, require a continuous or dichotomous health indicator (Wagstaff and van Doorslaer, 1994). Previous studies have aimed at resolving this situation, whereby a widely available categorical and ordinal variable is needed for the analysis of health inequalities. The result is a set of approaches that impose a cardinal scale on SAH, thereby providing various transformations for inequality analysis.

An early approach to inequality analysis that relies on the empirical distribution of SAH was proposed by Wagstaff and van Doorslaer (1994). This approach was proposed as an alternative to the common practice of comparing percentages of respondents in each category or dichotomizing the ordinal variable. It is commonly observed that the distribution of responses in SAH tends to be concentrated around the good health categories, while displaying a relatively thin tail for the categories corresponding to fair and poor health. Considering the skewness of the distribution, the approach of Wagstaff and van Doorslaer (1994) assumes that underlying the SAH responses is a latent, continuous health variable with a standard lognormal distribution.

A procedure to cardinalize the variable is to score the SAH categories as the expected values within each of the intervals defined by the cut points corresponding to the inverse lognormal distribution[2]. For instance, reporting a health level *h*, the health status of the individual is classified by the average value of the latent health variable between the thresholds (or boundaries)) λ_h and λ_{h+1}.The inequality health index derived from this method

[2] "The choice of an inverse or a standard lognormal distribution is explained by the skewness of the distribution. If this skewness is observed on the right (respectively left) then an inverse (a standard) lognormal would be preferred". (Tubeuf and Perorrini, 2008)

seems to be quite similar to other cardinalization procedures[3], as indicated by Gerdtham et al. (2000). The latent variable obtained in this way, however, is not bounded within the range [0,1] presupposed for a health utility score scale. Furthermore, the lognormal approach assumes the same distribution of health regardless the population considered, which is clearly a fairly strong assumption (van Doorslaer and Jones, 2002) and may result in a biased concentration index (Tubeuf and Perronnin, 2008 and Wagstaff et.al., 2001).

In order to account for variations in responses between population segments, other approaches have resorted to the use of multivariate techniques. Such is the case of the method proposed by Van Doorslaer and Jones (2003), which consists of modeling SAH using interval regression when the SAH boundaries are known. Unfortunately, these boundaries are not typically known *a priori*, and therefore the authors used a Canadian distribution of continuous health measure, namely Health Utility Index[4] (HUI Mark III), and scored the SAH intervals assuming a “stable mapping” between HUI and the latent health variable behind the subjective SAH. This implies that the *q*th quantile of the distribution of HUI corresponds to the *q*th quantile of the SAH variable. In this way, interval regression provides an estimate (conditional on socioeconomic factors) of each individual’s level of health utility preserving the HUI scale. Furthermore, importantly, the regression model provides estimated coefficients that allow for identification of the contribution of various factors, and an estimate of the variance of the error term (van Doorslaer and Jones, 2002).

From an implementation perspective, a difficulty is that many surveys do not include both SAH and a generic health measure to map boundaries between classes. It thus becomes necessary to assume that the empirical distribution function of the HUI scores, or some other generic health measure, such as EQ-5D[5] or SF-36, hold not only for their own samples, but also for other regions or even countries. Even so, this method has been widely applied in many countries, mainly based on Canadian HUI scores. For example, HUI Mark III scores have been employed to compare health inequality across different countries in Europe (Van Doorslaer and Koolman, 2004), to analyze regional health disparities in Spain (Gomez and Nicolas, 2004), and health index inequality in Jamaica (O’Donnell et. al., 2007).

---

[3] Comparable to those obtained using rating scale method, time trade off method (Gerdtham et al, 1999)
[4] Sampled by Canadian National Population Health Survey (NPHS 1994/1995)
[5] EQ-5D represents a generic measure of health developed by the EuroQol Group; SF-36 generic health indicator contructed by Medical Outcome Study

A more direct approach to model SAH in a multivariate framework is to use the ordered probit model. The ordered probit is an example of a model with a latent variable of the following form:

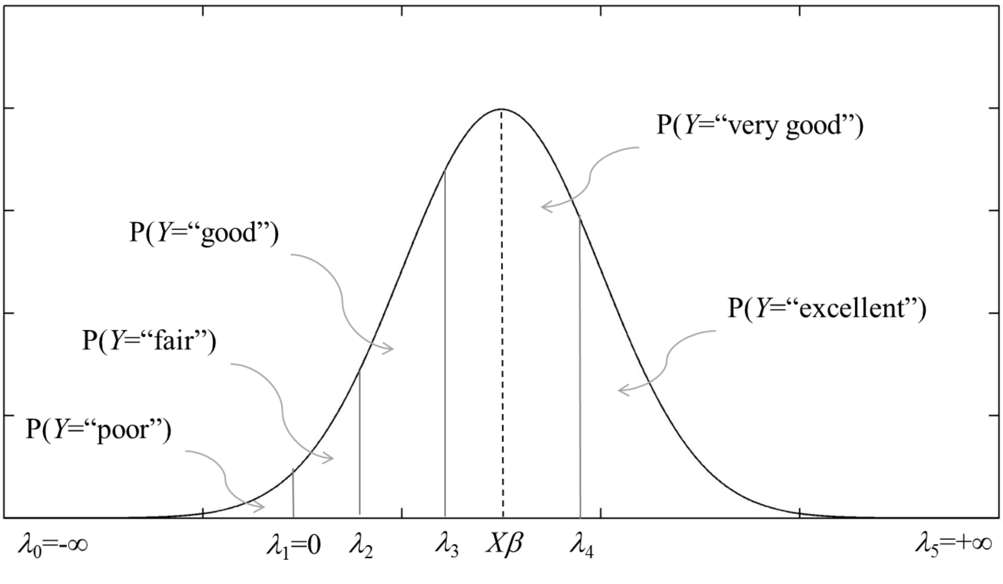
$$y_t^* = X_t \beta + \varepsilon_t \qquad (1)$$

In this model, a latent variable y_t^* is assumed to be a stochastic, linear-in-parameters function of covariates X_t. The random term ε_t is assumed to be identically and independently distributed, following the standard normal distribution. The latent variable is not observable. Instead, what is observed is an ordinal categorical variable Y_t, with level k (k=1,...,K), which is assumed to be determined as follows:

$$Y_t = k \text{ if } \lambda_{k-1} < y_t^* < \lambda_k \qquad (2)$$

Where the λ's are estimable threshold values that mark the boundary between ordinal classes. For instance, in the case of SAH, K=5, and the levels are k=1="poor", k=2="fair", k=3="good", k=4="very good", k=5="excellent"[6]. The latent variable can be related to the probability of a response being in certain category. This is simply the area under the curve, as illustrated in Figure 1.1:

Figure 1.1 - Latent variable and ordinal probit probabilities (area under the curve). Note that for identification λ_1=0. Other thresholds are estimated.



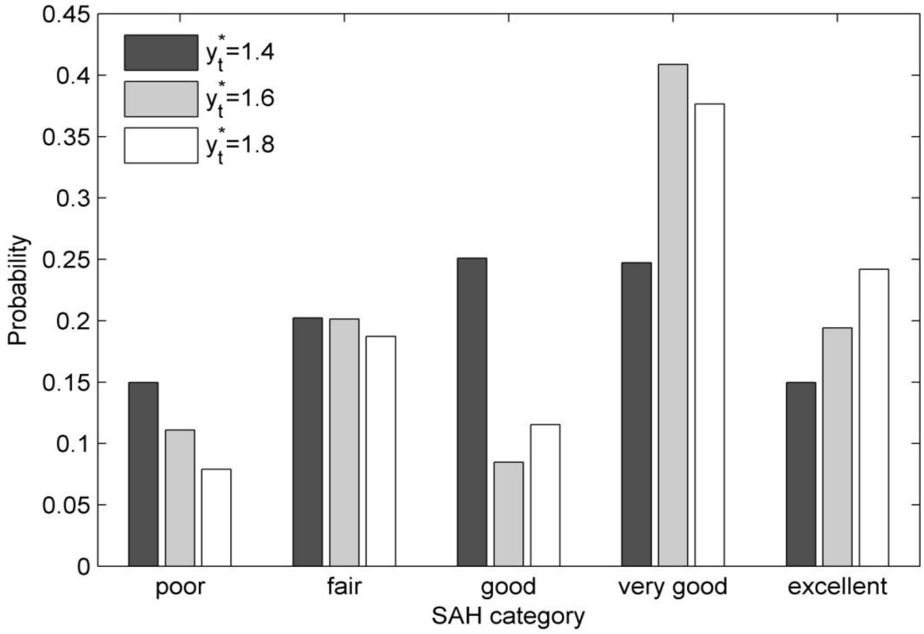
Source: Self Elaborated

[6] The categories are characterized from "poor" to "excellent" in many countries, as United States and Canada. However in Brazil, the categories ranges from "very poor" to very good". It doesn´t change the explanation´s logic of this section.

An attractive feature of the ordinal probit model is that it does not require prior knowledge of the boundaries between classes, since the thresholds are estimable from the data. Inequality analysis is implemented by using the estimated (linear and continuous) value of the latent variable $\hat{y}_t^* = X_t\beta$ to represent individual health status, perhaps with ex-post rescaling to ensure that the latent variable is bounded in the interval between 0 and 1 (van Doorslaer and Jones, 2002). Under such rescaling, 0 represents the worst health status and 1 the best – the rescaled variables are interpreted as "quality weights" or utility proxies. This approach has been adopted in multiple studies, including Lauridsen et al. (2004), van Doorslaer and Jones, 2003), Madden (2008).

A common feature of the methods discussed so far is that they enforce cardinality on SAH to obtain a continuous variable, while the underlying information is inherently categorical. This is problematic because, as noted by Apouey (2007), "enforcing cardinality is a 'supra-ordinal' assumption that changes the original properties of SAH". Perhaps the alteration in the original properties of SAH caused by 'supra-ordinality' is more easily illustrated by means of an example using the probit model. For instance, consider three values of the latent variable, say $\hat{y}_t^* = 1.4, 1.6, \text{and } 1.8$. Further, consider the following thresholds for the five classes: λ_0=-∞, λ_1=0, λ_2=0.9, λ_3=1.5, λ_4=2.8, and λ_5=+∞. Comparison of the linear-in-parameters latent variable gives the misleading impression that the differences in "utility proxies" are linear too, with the difference between 1.6 and 1.4 being the same as the difference between 1.8 and 1.6. However, these values are only meaningful in the context of the thresholds between categories, and the change in the resulting probabilities is not necessarily commensurate with the magnitude of the difference in the latent variable, as seen in Figure 1.2:

Figure 1.2 - Ordinal probit probabilities for different values of the latent variable.



Source: Self Elaborated

The problem is that the latent variable, while linear, is at the core of a non-linear model. Further, even if the latent variable is seen in a utility context, a change in the latent variable affects the utility of an individual expressing a certain opinion of the object of the question, and is not associated with the utility of the object itself. In other words, an increase in the latent variable may increase the probability of an individual having a better opinion of her health, but not necessarily commensurately better health. A relationship between the opinion and the object of the opinion cannot in fact be established, since all that is recorded is the opinion – and this in a categorical fashion, further compounding the problem of trying to impose cardinality in the ways discussed above.

The preceding discussion identifies the problems associated with trying to impose cardinality on what is essentially a categorical and ordinal variable. Instead, as an alternative, we suggest summarizing the information available in its native format, in such a way that the degree of inequality between individuals with different characteristics can be expressed by means of a continuous, directly comparable index. Its construction is composed into two main parts; the first relies on the Ordinal Model estimation and posterior probabilities prediction for specific population subgroups; and the second step, on the application of these probabilities into a procedure similar to Gini assessment for discrete variables in order to compute the index. We describe our proposed approach next.

1.2.2 AN INDEX OF INEQUALITY BASED ON THE ORDINAL MODEL

As denoted in the preview section, y_t* corresponds to a continuous and unobservable latent variable in an ordinal model. Thereby, our starting point to introduce the novel approach relies on assuming that the latent general health condition of the individuals, represented by y*, underlies the observable self reported health (SAH):

$$SAH_t = k \text{ if } \lambda_{k-1} < y_t^* < \lambda_k, \quad (k=1,...,5) \tag{3}$$

With $y_t^* = X_t\beta + \varepsilon_t$, ε_t~N(0 , 1) Λ t = 1…n

$$y_t^* = X_t\beta + \varepsilon_t \tag{4}$$

Where X_t is a vector of socioeconomic and demographic variables (described in the following section) and λ_k are the already denoted cut-off points of each category of self-reported health.

In this context, the vector of control variables assumes an important role once allows us to taking into account the individual-level variation in SAH responses. Hence, the predicted probabilities of each SAH category are computed conditional on the vector of individual characteristics:

$$P_{tk}(SAH_t = k)=P(\lambda_{k-1} < y_t^* < \lambda_k/x) \quad (k=1,...,5) \tag{5}$$

Substituting from (4)

$$= P(\lambda_{k-1} < X_t\beta + \varepsilon_t < \lambda_k/X_t) \tag{6}$$

$$= P(\lambda_{k-1} - X_t\beta < \varepsilon_t < \lambda_k - X_t\beta \;/\; X_t) \tag{7}$$

$$= \phi(\lambda_k - X_t\beta) - \phi(\lambda_{k-1} - X_t\beta) \tag{8}$$

Where ϕ is the standard normal distribution function. To estimate the whole model, the log-likelihood is employed and takes the following form:

$$LnL = \sum_t\sum_k Z_{tk} \ln P_{tk} \tag{9}$$

Defining Z_{tk} as the indicator variable that assumes value 1 if y_t=k and 0 otherwise.

The maximization of log-likelihood function gives us the coefficients β, the cut-off values and also the probability of being into a particular category “k” of SAH.

1.2.3 New Approach to Assess Gini Coefficient in Ordinal Contexts: The "Index-D"

The index proposed in the present paper, called Index-D, set itself up as a new approach to build a health inequality index through an alternative application of the already traditional Gini coefficient[7]. Thereby, we employ the probabilities predicted from the ordinal model described above, as our outcomes at a Gini equation set for discrete probability functions. As a result, we assess an inequality index, adapted for ordinal context, in a probability-basis.

Formally speaking, let us consider two hypothetical vectors of predicted probabilities corresponding to two different socioeconomic profiles:

$p_{hk}(SAH = k|profile\ h) = (\hat{p}_{h1} \dots \hat{p}_{hK})$
$p_{wk}(SAH = k|profile\ w) = (\hat{p}_{w1} \dots \hat{p}_{wK})$

$$\sum_{k=1}^{K} \hat{p}_{hk} = 1; \quad (10)$$

$$\sum_{k=1}^{K} \hat{p}_{wk} = 1 \quad (11)$$

Where p_{hk} and p_{wk} are respectively the set of predicted probabilities of the first and second vectors corresponding to each SAH category "k" ($k = 1 \dots K$). Given this, we are able to compute the two profiles´ joint probabilities functions, where *j* represents the number of possible "steps" along SAH outcomes, as we will explain onwards:

$$i. P_{kj} = p_{hk}p_{wk} \text{ for } j = 0 \quad (12)$$

$$ii. P_{kj} = P_{hk}\,(P_{wk+j} + P_{wk-j}) \text{ for } j > 0 \quad (13)$$

We now generate SP_j, that represents the joint probabilities´ sum of each step "*j*"; or our discrete probability function, corresponding to the discrete random variable (y_i):

$$f(y_i) \equiv SP_j = \sum_{k=1}^{K} P_{kj}; \text{ for } i = 1 \dots n; \ k = 1 \dots K \text{ ; and } j = i - 1 \dots n - 1 \quad (14)$$

Satisfying the following properties:

$$0 \leq f(y_i) \leq 1 \quad \text{and} \quad \sum_{k=1}^{K} f(y_i) = F(y_i) = 1 \quad (15)$$

For our ordinal variable SAH (n = 5), we have:

$k = 1 \dots 5$; and $j = 0 \dots 4$

[7] For details about Gini Coefficient, see (Bellù, 2006)

When $j = 0$ (zero steps along SAH range), it corresponds to the multiplication of the probabilities related to the same health category (k) of both vectors under consideration, which means the joint probability of this two groups assessing exactly the same health category, as demonstrated in equation (12). The procedure is analogous for the remaining possible steps:

Table 1.1 - Assessing the sum of joint probabilities for profiles *h* and *w* according to the number of "steps" j.

nº of steps (j)	Description	n° of outcomes $2(\mathbf{k}-\mathbf{j})$	Sum of the joint probabilities (SP_j)
1	joint probabilities of being one category distant (forth and back) of each other	8	$p_{w1} * p_{h2} + p_{w2} * p_{h1} + p_{w2} * p_{h3} + p_{w3} * p_{h2} + p_{w3} * p_{h4} + p_{h3} * p_{w4} * + p_{h4} * p_{w5} + p_{h5} * p_{w4}$
2	joint probabilities of being two categories distant (forth and back) of each other	6	$p_{w1} * p_{h3} + p_{w3} * p_{h1} + p_{w2} * p_{h4} + p_{w4} * p_{h2} + p_{w3} * p_{h5} + p_{h3} * p_{w5}$
3	joint probabilities of being three categories distant (forth and back) of each other	4	$p_{w1} * p_{h4} + p_{w4} * p_{h1} + p_{w2} * p_{h5} + p_{w5} * p_{h2}$
4	joint probabilities of being four categories distant (forth and back) of each other	2	$p_{w1} * p_{h5} + \cdot\, p_{w5} * p_{h1}$

Source: Self Elaborated

Intuitively, these results are indentifying the probability of moving from a health category k to a category k' whereas we move from profile h to profile w.

Defining φ_i as the steps´ frequency, ($\varphi_i = SP_j j$), we generate the cumulative distribution function, S_i. Where $S_n = \sum_{i=1}^{n} \varphi_i$ and $S_0 = 0$.

Once computed S_i, we are able to apply the Gini formula suitable for discrete probability functions, and estimate our Index-D:

$$G_D \equiv I_D = \sum_{i=1}^{n} f(y_i)(S_{i-1} + S_i)/S_n \qquad (17)$$

Let us illustrate the profiles regarded above by means of both numerical and graphical example in order to clarify the index-D interpretation:

$p_{tk}(SAH = k|profile\ h) = (\hat{p}_{h1};\ \hat{p}_{h2};\hat{p}_{h3};\hat{p}_{h4};\ \hat{p}_{h5})$ = (0.01; 0.09; 0.1; 0.3; 0.5)

$p_{tk}(SAH = k|profile\ w)$ = $(\hat{p}_{w1};\ \hat{p}_{w2};\hat{p}_{w3};\hat{p}_{w4};\ \hat{p}_{w5})$ = (0.07; 0.15; 0.2; 0.15; 0.43)

Notice that the first vector displays higher probabilities in favor of the top health categories than the profile w. Applying to them the procedure just described above, enables us to verify how much "unequal" these two set of predicted probabilities are from each other:

Table 1.2 - The discrete probability function of steps j and the respective accumulated frequency of profiles *h* and *w*

nº of steps (j)	0	1	2	3	4
SP_j	0.2942	0.3198	0.2105	0.1362	0.0393
S_i	0	0.3198	0.7408	1.1494	1.3066

Source: Self Elaborated

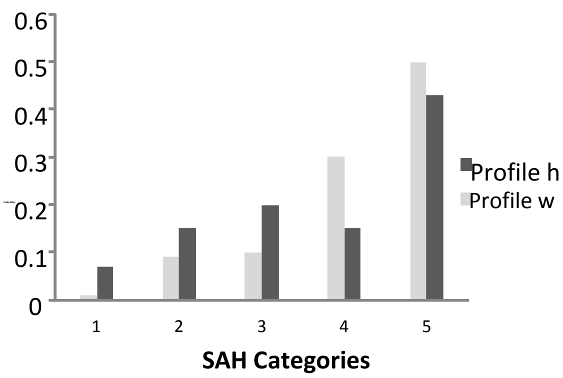
Employing these outcomes into the formula (17) gives us the level of inequality for this example: index-D = 0.52004.

Additionally, we consider a third vector representing a new socioeconomic group:

$p_{tk}(SAH = k|profile\ z)$ = (0.04; 0.06; 0.1; 0.2; 0.6)

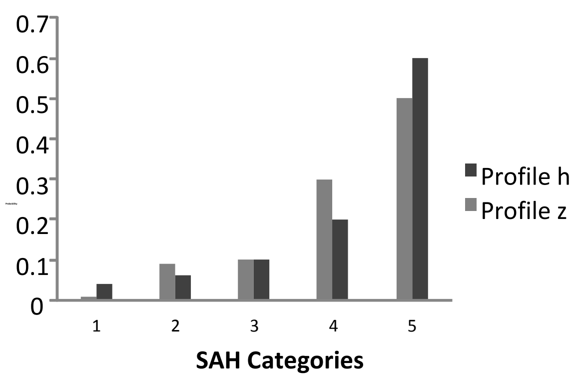
It is noteworthy that the predicted probabilities of the new profile *z* tend to be more similar to the profile *h* than to profile *w*, and as a result, the index-D is now equal to 0.4587. The Figures 1.3,1.4,1.5 and 1.6 illustrate this:

Figure 1.3 - Predicted Probabilities of Profiles *h* and *w*



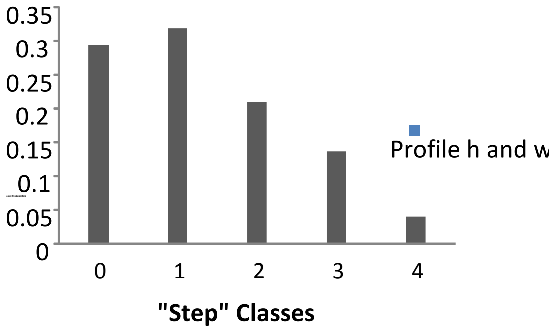
Source: Self Elaborated

Figure 1.4 - Predicted Probabilities of Profiles *h* and *z*



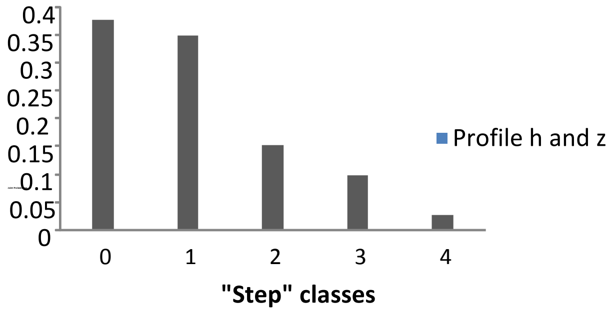
Source: Self Elaborated

Figure 1.5 – Discrete Probability distribution of steps *j* between *h* and *w* profiles



Source: Self Elaborated

Figure 1.6 - Discrete Probability distribution of steps *j* between *h* and *z* profiles



Source: Self Elaborated

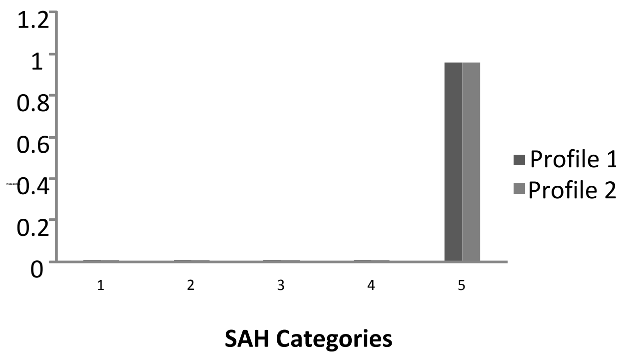
Figures 1.3 and Figure 1.4 display only what we previously reported through the vectors: the predicted probabilities of answering one of the five SAH categories. Using such information, we can compute, as explained above, the probability of moving from a SAH category k to category k' as long as we compare profile *h* with profile *w*; and profile *h* with *z*. As a result, we have the discrete probability distribution of the *j* steps, as reported in Figures 1.5 and Figure 1.6. It is noteworthy that the probability of steps 2, 3 and 4 are slightly higher between the groups *h* and *w* (Figure 1.5) than in Figure 1.6; or on the contrary, the step zero is somewhat lower in Figure 1.5.

Intuitively, it means that comparing the propensity of answering a determined health category, the first two profiles hold a slight tendency of being more spread out along the range of SAH outcomes, and as a result, there exists a tendency of larger steps along SAH range. On the other hand, the groups *h* and *z* are likely to answer more similarly, i.e., there is a higher probability of moving only one category (step 1) or simply do not move (step 0). Consequently, a lower degree of inequality is estimated for this second example.

Furthermore, the index can also assess the inequality intra-class. In other words, it is possible to analyze how equal (unequal) is a distribution in reported health within a specific group that one has interest to study. Using the same examples, we can verify that the profile *h* has a within index equal to 0.4823; whereas, the second and third groups has respectively, I_{Dw} =0.5383 and I_{Dz}= 0.4218. In this context, we can implement the index in a practical and direct way without any cardinality assumption neither external scores.

In the classical income distribution hypothesis, the Gini coefficient is null if each individual in the population holds the same percentage of income; in contrast, if only one individual has the total percentage of income, the Gini will assume its maximum value corresponding to 1 (Bellu,2006). In our framework, as we have categorical variables representing subjective health, the index-D tends to a zero value when all individuals from two population groups (profile 1 and 2) rank their perceived health identically (Figure 1.7). On the other hand, if for instance, all individuals from one group (profile 3) assess very poor health and all individuals from the other group (profile 4) respond very good health (Figure 1.8), the index-D tends to one, indicating perfect inequality between groups, as we can see in the Figures 1.7 and 1.8:

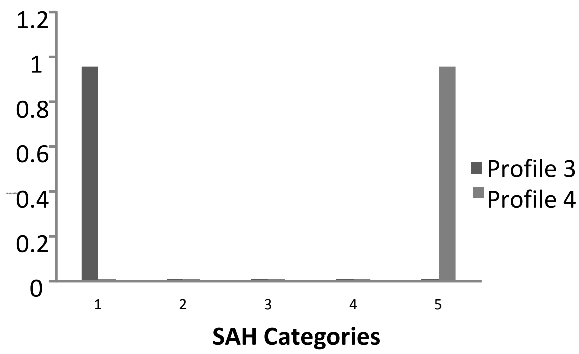
Figure 1.7 - A case of "*quasi-perfect*" equality in Self Assessed Health



Index-D= 0.0584

Source: Self Elaborated

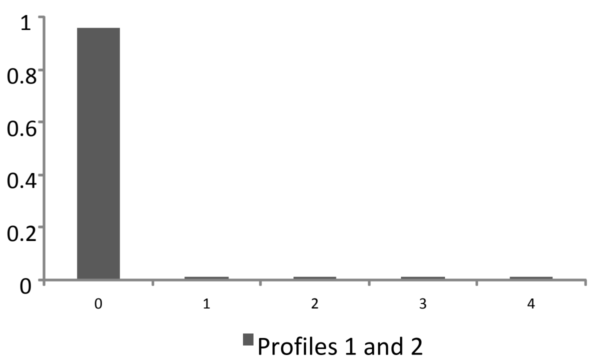
Figure 1.8 - A case of "*quasi-perfect*" inequality in Self Assessed Health



Index-D=0.9519

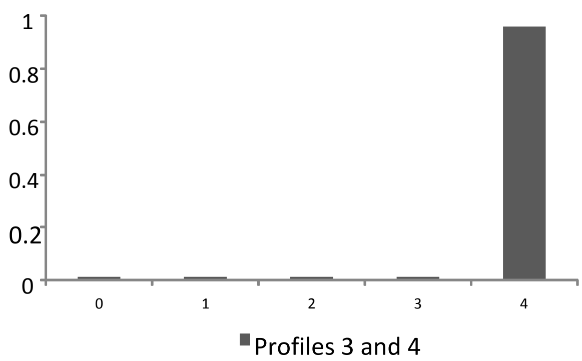
Source: Self Elaborated

Figure 1.9 - Discrete Probability distribution of steps *j* for "*quasi-perfect*" equality in SAH categories



Source: Self Elaborated

Figure 1.10 - Discrete Probability distribution of steps *j* for "*quasi-perfect*" inequality between in SAH categories



Source: Self Elaborated

It is important to note that a challenge that characterizes indices of inequality based on categorical variables relies on ranging its value within the 0-1 interval. This desired property known as “*normalization axiom”* holds for the Gini Coefficient as well as for our approach.

Figure 1.9 shows that the probability of do not move any step along SAH range is practically one, since both profiles tend to respond the same health category. It corresponds, as a result, to a“quasi-perfect” health equality. Conversely, the “quasi-perfect” inequality arises in Figure 1.10, where profiles 3 and 4 perceive their health in opposite directions, and the probability of 4 steps along SAH categories tends to one.

1.3 CONTEXT AND DATA

The empirical application of the current paper inserts itself in the context of a country marked by substantial demographic and socioeconomic inequalities. Brazil is historically classified as one of the countries with the greatest income inequality in the world – in 2010, the highest 10% salaries among the population held 44,5% of total income, whereas the lowest 10% detained 1,1% of the national income. Despite of the recent optimistic statistics – the Gini index decreased from 0.548 in 2003 to 0.496 in 2009 – the country still presents deep and persistent socioeconomic disparities, as inequality in schooling attainment, housing conditions and the access (and quality of) to the public services that manifests differently across regions, states, center-periphery, and urban-rural areas.

In this context, Brazil represents a rich setting for inequality analysis, since its population holds high variation of socioeconomic aspects both within and between socio-demographic groups and across regions. Surprisingly, the health inequality analysis in a socioeconomic context has been a scarce subject in previews studies in the country. It is important to emphasize either that Brazil holds a public health system (Sistema único de Saúde -SUS) that focus on universal access on a decentralized basis. However, coverage all demand (74% of the population)[8] with good quality services is an aim far from being reached. Furthermore, there is still lack of public sanitation in many urban and rural areas, particularly in North and Northeast regions.

1.3.1 DATA AND DESCRIPTIVE ANALYSIS

We provide an empirical illustration of the index-D using Brazilian microdata from the 1998, 2003 and 2008 National Household Survey (Pesquisa Nacional por Amostra de Domicílio-PNAD) in order to show the usefulness of this measure. The PNAD database is an annual[9] national survey of households which collects an extensive set of personal, demographic and socioeconomic variables, as labor participation, income (values and source), schooling attendance, family composition, migration and infrastructure conditions of the households. In each year, a range of supplementary questions is added to the survey, addressing different social subjects. Particularly in 1998, 2003 and 2008, this supplement

[8] According to PNAD 2008
[9] Except when the national survey (census) is collected and in 1994 for budget constraint

relied on health information, which include, besides our key variable self assessed health, using continuous medication[10]; health insurance; chronic diseases (clinical diagnosis); the frequency that the person utilized the health public service in the last year and its reasons; functional capabilities, as mobility and self care.

We kept only people aging 16 to 65 years old in order to reduce the potential reporting bias due to age. Also, it represents only people potentially included in labor market. The sampling scheme is self-weighted to be population representative. Our final sample for the years 1998, 2003 and 2008, contemplates respectively, 206,070; 248,114; and 259,118 respondents. Table 1.3 provides the frequency and the percentage of each SAH category along with its mean and standard deviation for the three years under investigation.

Table 1.3 – Descriptive Statistic of Self-Assessed Health for 1998, 2003 and 2008

Self Assessed Health (SAH)	1998		2003		2008	
	Mean	**St.Dev.**	**Mean**	**St.Dev.**	**Mean**	**St.Dev.**
	3.966	**0.793**	**3.95**	**0.7686**	**3.912**	**0.7702**
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**
Very Poor	1,345	0.65	1,436	0.58	1,872	0.72
Poor	6,522	3.17	7,356	2.96	8,078	3.12
Fair	42,127	20.45	51,112	20.6	54,795	21.15
Good	107,104	51.99	135,135	54.46	142,766	55.1
Very Good	48,920	23.75	53,075	21.39	51,607	19.92
Total	**206,018**	**100**	**248,114**	**100**	**259.118**	**100**

Source: PNAD 1998, 2003 and 2008

Following the majority surveys containing SAH, most of the population reported good or very good health, conferring the skewness shape of the variable. Considering these three years, the proportion of respondents in each category followed the same pattern, though it is noticeable that 2003 has a slight trend of being more optimistic in perceived health. Also, 2008 presents the highest frequency of very poor health and the lowest of very good health. Still, 1998 holds the highest mean, followed by 2003 and 2008.

One issue is important to bear in mind when it comes to analyze such numbers. The answers may have been influenced by the circumstances and happenings related to the year that the survey occurred. Since being subjective, it is susceptible, amongst other factors, to individual's expectations to the current economic and social environment in which they live in, as for example, labor market expectations.

[10] Only for 2003 and 2008

Table 1.4 presents the summary of all socioeconomic variables in the Ordered Probit Model. For ease of exposition, and also because we restrict the analysis only for the 2008`s database, we display and discuss the following descriptive statistics only for this referred year. The results for 1998 and 2003 are attached.

Table 1.4 - Independent Variables of the Probit Model: Descriptive Statistics for the year 2008

VARIABLES	*code*	Description	nº of obs.	Mean	SD
Gender	*gender*	Men=1	259118	0.482	0.499
Age	*age*	The age variable ranges from 16 to 65 years old	259118	36.342	13.46
Race	*race*	Dummy representing white=1 and non-white=0	258070	0.456	0.498
lnincome	*ln_pcincome*	Per capita household´ income in log scale. Min.= 0.847 and Max=11.918	253379	5.944	1.005
Years of schooling	*schooling*	Number of years of schooling. Ranges from 1 year or less to 16 years or more	257847	8.858	4.32
Couple with children	*married_child*	Couple has one child or more =1 (Dummy)	259118	0.611	0.487
Couple without children	*married_nochild*	Couple has no child = 1 (Dummy)	259118	0.132	0.338
Single mother	*single mother*	Mother with both children younger and/or older than 14 years old (Dummy)	259118	0.160	0.367
Single (no child)	*single*	The individual is not married neither has child = 1 (Dummy)	259118	0.095	0.293
Worked in the week of reference	*employed*	The individual is economically active and worked in the week of reference = 1 (Dummy)	259118	0.662	0.472
Did not work in the week of reference	*unemployed*	The individual is economically active but didn´t work in the week of reference = 1 (Dummy)	259118	0.083	0.276
Health insurance	*health_ins*	The individual holds private health insurance = 1 (Dummy)	259118	0.264	0.441
urban_rural areas	*urb*	Urban area=1; rural=0	259118	0.853	0.353
Metropolitan Region	*reg_metrop*	Metropolitan region=1 ; non metropolitan region = 0	259118	0.381	0.485

Source: PNAD 2008

In order to take into account the individual level variation in self reported health, we selected a set of socioeconomic and demographic factors that have been extensively reported in previews literature (Dahlgren and Whitehead's Social Determinants of Health Rainbow (1991) APUD SACOSS[11], 2008; Doorslaer and Jones, 2002) as important determinants of both variations in general perception of health and reporting behavior bias[12].

The first set of variables concerns about demographical information of respondents, and represents an important control, since gender and age are known to have a strong effect on the divergence between the "true" and reported health. For example, the difference in perception between 'poor' and 'very poor health' is smaller for older people than for the young (Groot, 2000). The gap in health perception between man and woman is also remarkable. Man usually report better health, despite the lower life expectancy.

The income class, in turn, has been strongly associated with a range of health outcomes. Our measure is reported in terms of household per capita; it is continuous and was transformed into natural logarithm to have its scale reduced. The years of schooling along with marital status, are also well documented as impacting variables for perceived health (SACOSS, 2008). We thus indentify that, for our sample, the population has attended the school per 8.8 years on average; and 61% are couple with children. Regarding the work status, 66% worked in the week of reference, whereas only 8% did not[13]. Furthermore, health may be significantly affected by accessing health services (Wagstaff, 2001). In this sense, holding private health insurance is an important control variable, especially for being a status unequally distributed across socioeconomic groups in the country. In fact, only 26.4% of the population holds a private health service.

Finally, the role of geographical dimension has been object of interest in many researches (Wagstaff , 2001; Habibov et. al. 2011) and requires special attention on health inequality analysis. We consider thus, the dummies representing urban versus rural area – where 85% of the population are in urban cities; Brazilian metropolitan regions and also the 26 federative states.

---

[11] South Australian Council of Social Service

[12] It is important to note, however, that not all health relevant social aspects were incorporated in our model, as for example, variables related to childhood (including prenatal period) and parental schooling, which are rarely available information in wide-ranging surveys.

[13] There is still the omitted group representing those who are economically inactive

1.4 RESULTS

The table 1.5 reports the coefficients from the Ordered Probit Model on Self-Assessed Health and the socioeconomic variables detailed above for 2008. Such results are the first step toward the index interpretation. Once we employ such variables to assess the inequality indices and compare different socioeconomic profiles, it is essential to indentify previously the direction and degree of correlation between the variables of interest and the self reported health of Brazilian population. Most coefficients presented the expected signal with high statistic significance.

Table 1.5 - Results of the Ordered Probit Model - 2008

Dependent Variable: Self Assessed Health

Variables	**Coeff.**	**S.D**
gender	0.1113	0.0002
age	-0.0304	0.00004
age2	0.00007	6.18E-07
race	0.0728	0.0002
schooling	0.0266	0.0001
esc2	0.0008	5.760E-06
ln_pcincome	0.15027	0.0001
employed	0.1826	0.0002
unemployed	-0.0124	0.0004
married_nochild[1]	-0.075	0.0004
married_child	0.017	0.0003
mother	-0.0148	0.0004
health_ins	0.0662	0.0002
reg_metrop	0.0199	0.0002
urb	-0.0438	0.0003
cut1	-2.358	0.0015
cut2	-1.603	0.0015
cut3	-0.365	0.0015
cut4	1.338	0.0015
n	251.194	
Log-Pseudolik.	-258225.76	
Wald chi2	43435.6	

Source: PNAD 2008

Note: 1.The marital status indicators are related to the omitted variable: "*single status*"; 2.The states' dummies were omitted

In agreement with the literature, men are more likely to report better health than woman, and older people are more willing to answer poorer health than the younger. Likewise, the better-off are highly prone to answer good or very good health than the lower income class. This evidence has been widely reported, and we also focus on that information, especially because Brazil inserts itself in a context with profound income disproportion, as briefly commented above. Thus, it seems relevant to analyze the self perceived health exploring the huge income variability across the country.

Likewise, the coefficient and significance of work status suggest the positive role of being employed. Concerning about marital status, it is notable that single mothers own lower chances of saying better health.

Despite of the low coefficient value, respondents who hold a private insurance health in Brazil are likely to say better health than those who need the health public service. Still, the metropolitan region´s residents are more prone to say better health than those from other localities.

1.4.1 INEQUALITY ANALYSIS USING INDEX-D: AN EMPIRICAL ILLUSTRATION

Once the Probit Model was estimated for the whole population, the index-D enables us to investigate SAH inequality for as many social dimensions as included in the model, and therefore, classify the population into reference groups, such as income class, gender, age, ethnicity, etc. or arrangements of these. We focus on this paper, on women’s reported health, and many reasons support this choice. In both developed and developing countries, women are more likely to report poorer health than men. There are international evidences showing that multidimensional aspects of life, such as social, cultural, economic and biological have a stronger effect in the health of women than in the men’s (Sen and Ostlin, 2007). In addition, they are more likely to be employed receiving lower wages and occupying informal positions, along with the burden of domestic work (Ministry of Health, Brazil, 2004).

In Brazil, women are the main users of the public health service (SUS). They use it not only for their own care, but also to follow their children and other relatives, such as the elderly (Ministry of Health, Brazil, 2004). Being aware of such evidences, the Brazilian government has been creating specific focused-on-women Programs and campaigns, as, for instance, the “National Plan for women”; the “National Program for Control of uterus cervix”;

and the “Rede Cegonha”[14] program. Along these lines, we assume the existence of socio-demographic groups amongst women, holding meaningful health differences. Thereby, the index-D represents a useful tool to provide empirical evidences still much scarce in the country about an issue of great importance for Government’s policies and programs conduction.

We draw, a priori, certain reference groups representing distinct social gradients on society, and apply the index-D in the sense of verifying in what extent the women´s perceived health may vary, as we change one favorable social characteristic to another from the bottom of society, as for instance, from the highest to lowest level of schooling. Also, the index enables us to analyze how unequal (equal) is the subjective health among women within a specific socio-demographic group (intra-class inequality) and if it presents relevant changing across regions and years. It is important to highlight that the index-D values in themselves say little; what enable us to interpret, and make it explicative, is when we compare them in different circumstances, time, space or combination of them.

The first dimension analyzed, focus on the social indicator most reported in literature of health inequality: the income class. The women were classified into household income per capita quintiles; i.e., from the 20% poorest to the 20% better-off; and into two distinct age groups: 16 to 37 and 49 to 65 years old represented by groups 1 and 2 respectively. The results are displayed in Figure 1.11, and provide the evidence of a decreasing inequality over time, for all five income classes.

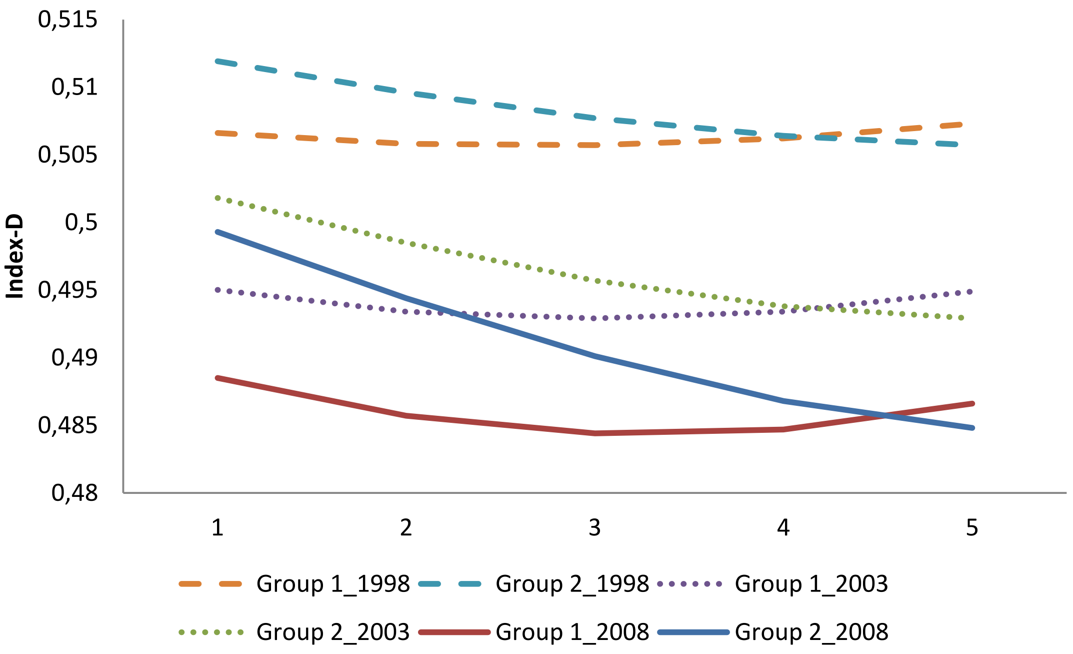
The inequality decreased, from 1998 to 2008, 1.72% and 2.03% on average among the old and young groups respectively[15]. It is worth noting that, if we look separately each quintile, such decrease was higher for the better-off, especially amongst the older group. As expected, group 2 perceived their own health more differently amongst them, and, as a result, held a higher index-D. However, such evidence is curiously reversed in the better-off quintile indicating a slightly higher inequality within the younger group.

[14] Aims to reduce maternal and neonatal mortality

[15] These values are in consonant with Gini´s literature. For example, the Gini coefficient decreased 4,6% between 2001 and 2005 (from 0,593 to 0,566) in Brazil, which represented one of the most accelerated decrease among 74 countries. (Barros et. al., 2007)

It is important to remember that the Gini index does not consider the income level; only how unequal is its distribution along the accumulated population. Analogously, the Index-D does not generate information about the average level of health; a younger group has, on average, a better health status than an older one, but not necessarily its index of inequality is lower, as we just accomplished from the given example. In this context, the inequality health measurement by itself does not provide all the necessary information to guide health policies. The population health´s mean level also matters significantly. In this sense, policymakers are prone to trade inequality with the health´s mean level, if the latter increases substantially (Kawachi et. al.,2002). Concerning this issue, we provided, attached to the index-D results, the predicted probabilities of reporting ‘good’ or ‘very good’ health for each group under consideration.

Figure 1.11 - Within Index-D for women aging 16 to 37 (Group 1) and 49 to 65 (Group 2) per income´s quintiles - 1998, 2003 and 2008



Souce: PNAD 1998,2003,2008.

Table 1.6 – Predicted probabilities of answering "good" or "very good" health for women aging 16 to 37 (Group 1) and 49 to 65 (Group 2) per income´s quintiles - 1998, 2003 and 2008

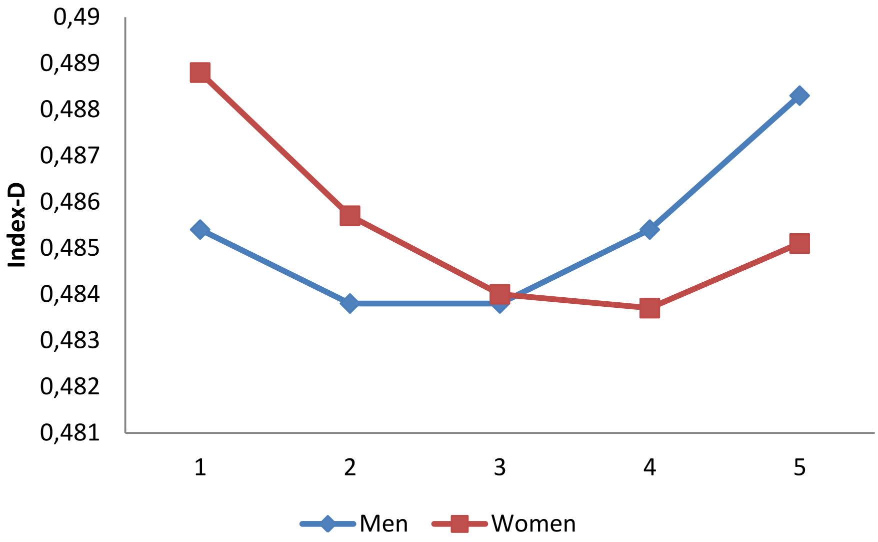
Group 1		**quintile 1**	**quintile 2**	**quintile 3**	**quintile 4**	**quintile 5**
Prob. (SAH>=4)	**2008**	0.7044	0.7381	0.7698	0.7992	0.8262
	2003	0.7169	0.7461	0.7736	0.7994	0.8234
	1998	0.7139	0.7422	0.769	0.794	0.8176
	Δ 2008-1998	*-0.95%*	*-0.41%*	*0.08%*	*0.52%*	*0.86%*
Group 2						
Prob. (SAH>=4)	**2008**	0.6149	0.6528	0.6891	0.7237	0.7562
	2003	0.6425	0.675	0.7062	0.736	0.7641
	1998	0.6279	0.6596	0.6903	0.7197	0.7477
	Δ 2008-1998	*-1.30%*	*-0.68%*	*-0.12%*	*0.40%*	*0.85%*

Souce: PNAD 1998,2003,2008.

It is noticed through Table 1.6, a greater bias on responding 'good health' amongst those with better financial condition, representing the existence of a clear social gradient – regardless the year and the age group. The difference between 2008 and 1998 also describes an interesting result: for the bottom quintiles, the likely of responding 'good health' is lower in 2008 than in 1998; whereas among the better-off, the opposite result arises.

If we enhance only the year of 2008 and estimate the *intra-class* Index-D regarding the whole population, but the genres separately, an interesting result occurs. As showed in Figure 1.12, women who belong to the first income quintiles group holds an index 0.34% higher than the men group. This same difference is seen on the richer quintile but conversely, it comes from a higher inequality in the men group. This does not mean, as elucidated before, better or worse health status, but only how spread out are the responses within these groups.

Figure 1.12 - Within Index-D per Income Quintile for Men and Women-2008



Source: PNAD 2008

Table 1.7 - Predicted Probabilities of answering "good" or "very good health" per Income Quintile for Men and Women

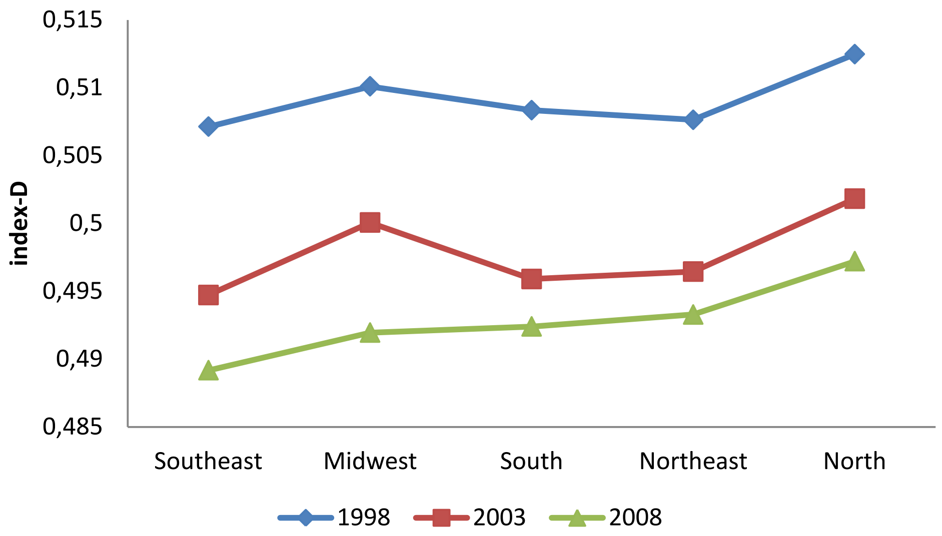
P(SAH>=4)	**Men**	0.734	0.765	0.794	0.821	0.845
	Women	0.697	0.729	0.761	0.79	0.817
$\hat{p}_{men} - \hat{p}_{women}$		3.73%	3.60%	3.30%	3.13%	2.80%

Source: PNAD 2008

In particular, the level of health status, represented here by the predicted probability of responding "good" or "very good'" health presents favorable results for men, consistently attested in previews literature. It is important to highlight, nevertheless, that this disparity between genres tend to decrease as they rise on the social scale. Such results may suggest that compared to men, the women are on average, more "sensible" or even more vulnerable to the disadvantage social circumstances. It is worth further investigations about this differentiability.

"The health and disease profile vary in time and space, accordingly to each region's degree of economic, social and human development" (Laurell, 1982). In this sense, the role of place and context of residence (such as neighborhoods, workplaces, regions, states) in reproducing health inequalities is a topic of increasing concern in literature (Kawachi et. al. 2002).The contextual heterogeneity is well traced in Brazil and represents a rich source of relevant information through huge social disparities between its regions. In this sense, we also investigate women´s SAH inequality across the five Brazilian regions. The next illustration brings the inequality between metropolitan versus non metropolitan areas considering only the women aged 49 to 65 years old without health insurance. The idea is to examine the differentials in reported health between an area that offers more health services accessibility – the metropolitan region – and, on the other side, smaller urban centers and rural area – where the latter is known by the scarce health and sanitation services. Furthermore, the profile considered is the age group that, on average, most needs health services[16]and in addition, its members are potential users of the public health service (SUS), once they do not have health insurance.The intuition behind this comparison relies on asking in what extent a woman without private health care, who lives in a metropolitan area, differs her perceived health from a women with similar profile, but lives in a smaller town or rural area.

Figure 1.13 - Index-D between Metropolitan versus Non-Metropolitan Areas for women aging 49 to 65 years old per Brazilian Regions – 1998, 2003, 2008



Souce: PNAD 1998,2003,2008

Table 1.8 - Predicted Probabilities of answering "good" or "very good health" between Metropolitan and Non-Metropolitan Areas for women aging 49 to 65 years old per Brazilian Regions – 1998, 2003, 2008

P(SAH)>=4	**2008**		**2003**		**1998**	
	rm	**non rm**	**rm**	**non rm**	**rm**	**non rm**
Northeast	0.6713	0.6629	0.715	0.7093	0.7022	0.729
North	0.6396	0.6308	0.654	0.6477	0.6167	0.6466
South	0.6789	0.6705	0.7234	0.7177	0.6845	0.7122
Southeast	0.7094	0.7013	0.7474	0.742	0.7196	0.7457
Midwest	0.6828	0.6745	0.6717	0.6655	0.652	0.6807

Souce: PNAD 1998,2003,2008

[16] Within the sample we are working:16-65 years old

Again, results from Figure 1.13 display the significant difference across the years investigated. This means that the women aging 49 to 65 years old from the metropolitan regions, perceived their health more differently than those living in other areas in 1998 than in 2008. In particular, southeast and midwest reflected the highest difference between 1998 and 2008 indices (1.8%;18.2%); whereas North and Northeast obtained the smaller ones – 1.53% and 1.44% – respectively. Moreover, North region presented, again, the highest index for the three years.

Looking the predicted probabilities at Table 1.8, 2003 and 2008 presented the expected results, i.e, the group from metropolitan region is prone to answer better health than those who live in smaller towns or rural areas.Surprisingly, for 1998, we verify a reversed result for all regions. Along with such an evidence, another outcome that draws attention is the lower probabilities of 2008 in comparison to 2003 and 1998; for the most results.

Deep explanations of these evidences are out of this paper's scope, but we need to bear in mind that Brazil has undergone significant social and economic transformation between the years of 1998 and 2008. Specifically in the health area, greater access to medical diagnostics for all income layers of the population leads to a greater information level and consequently, to possible alterations in the self perceived health. Besides that, increasing information through programs and/or public health campaigns drives the same population to health care units. Therefore, those people that, beforehand, had little or no access to health programs or medical services, started being instructed and more well-informed and, as a result, became more aware about their own "objective health".

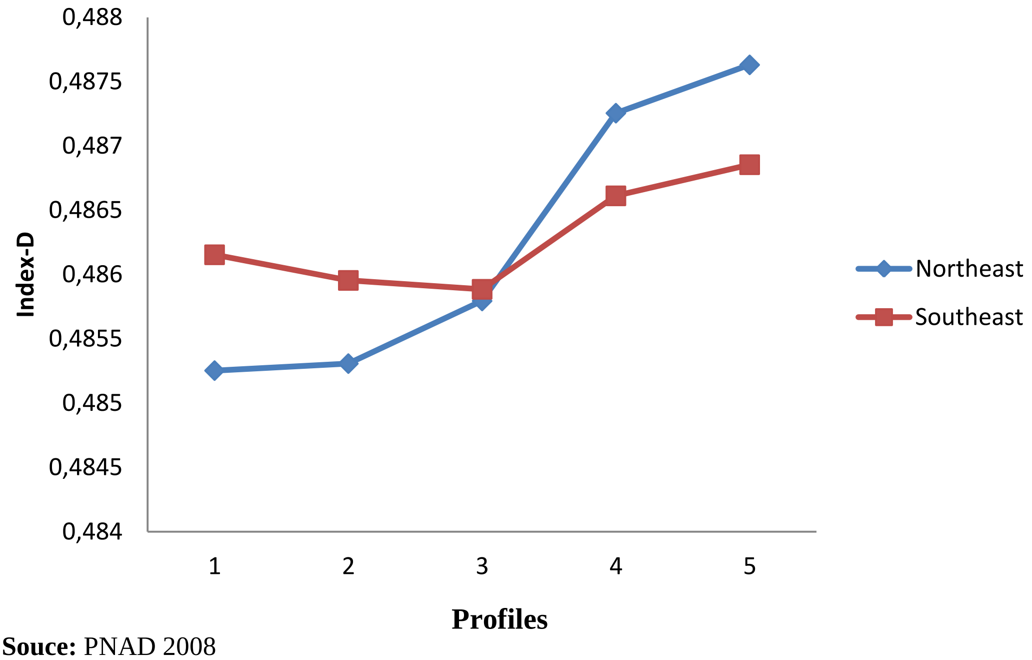
Furthermore, despite the rates of infant and matern mortality, malnutrition and dieseases that normally plague undeveloped coutries are decreasing in Brazil – the country has presented, in recent years, increasing rates of diseases tipically from developed world, such as cardiovascular and chronic degenerative . Such evidences might and should be investigated in subsequent studies in order to clarify how the self reported health in Brazil might reflect the new "health trend" on one side, and, on the other side, how much may be explained by the simple fact that the population is now accessing more often health diagnosis.

Back to index analysis, another analitical possibility of the Index-D relies on pinning a specific profile as reference and adding to this one disadvantaged socioeconomic characteristics. The Figure 1.14 displays this procedure for northeast and southeast regions;

both known by their historical disparity (in which the Northeast is lagging behind regarding socioeconomic development). The analysis starts with the intra-class (within) Index-D for single mother with child less than 14 years old, representing a group naturally more vulnerable in society[17]. The following indices are computed comparing the new profiles with the single mother vector (reference group) in order to verify in what extent the index alters as we add new characteristics to this group (see table 9 bellow).

In regards of the first within index-D, southeast region presents a slightly higher inequality (0.9%), as showed by Figure 1.14. However, as we agregate the new characteristics, Northeast pulls ahead. The main result relies on the increase of the index (0.14% and 0.07% for Northeast and Southeast, repectively) as we compare the reference group (single mother on average) with those non-white single mother with low schooling and low income. In addition, if we consider only those without health insurance (profile 5) the index still slightly rises.

Figure 1.14 - Index-D for different profiles of single mothers. Northeast and Southeast- 2008



Souce: PNAD 2008

[17] Approximately 2.8% of Brazilian populations are single mothers of children equal or younger than 14 years old, which represents 3.508.700 women.

Table 1.9 - Predicted Probabilities of answering "good" or "very good health" according to different single mothers´ Profiles. Northeast and Southeast - 2008

Pr(SAH>=4)		**SE**	**NE**
Profile 1	single mother	0.808	0.778
Profile 2	non-white single mother	0.801	0.768
Profile 3	non-white single mother with low schooling	0.78	0.748
Profile 4	non-white single mother with low schooling and low income	0.744	0.717
Profile 5	non-white single mother with low schooling; low income; without health insurance	0.742	0.711

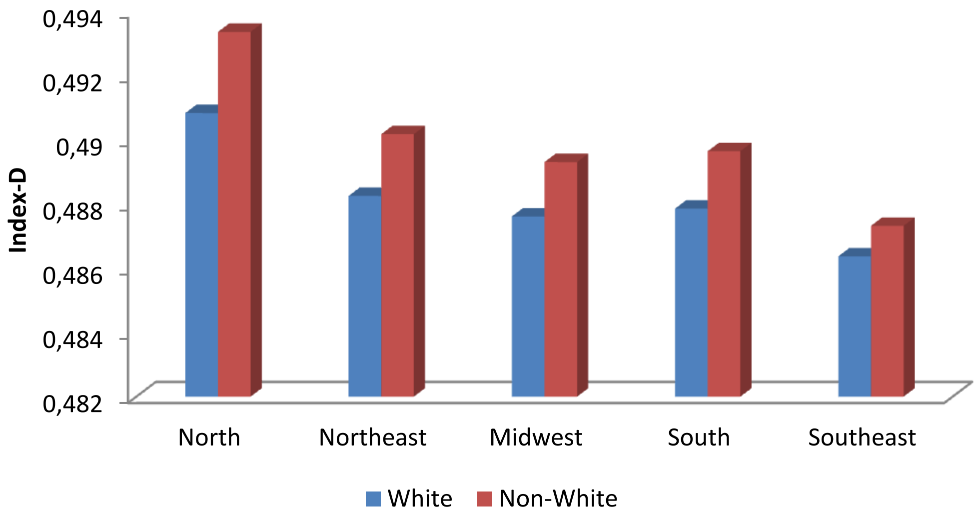
Souce: PNAD 2008

Considering the predicted probabilities, we can see through Table 1.9 the probability of anwering a better health presents a declinant trend across the profiles for both regions, and single mothers from Northeast are less likely to say good health than their similars in Southeast, for any profile. Put in another way, as the single mothers "go down" in the social scale, the inequality increases (specially in Northeast) while the health status presents a declinant tendency in direction to poorer health. These results are expected *a priori*, and reinforce the hypothesis of the perverse influence of social deprivation on self reported health – at least for women.

Let us now consider only the poorest women of the 2008 year (who belong the first and second income quintiles) and address the role of schooling attendance. The poorest women who attended school only for five years or less, are compared with those who had more than eleven years of schooling[18]across regions and race group (white; non-white). Looking at Figure 1.15, we note that the difference in the assessed health among those more and less educated varies according to the region's development grade. Again, the North and Southeast hold, respectively, the highest and lowest inequality degree. Furthermore, the non-white group compared to the white one, presents a slightly greater inequality in all regions, being this difference higher in the North and lower in the South (0.25% and 0.09%, respectively).

[18] The proportion of women from income quintiles 1 and 2 who hold equal or less than 5 year of schooling is 38.7%, while the women who attended more than 12 years is 2.7%

Figure 1.15 - Index-D of the poorest women with High and Low Schooling, per Region and Race - 2008



Souce: PNAD 2008

Table 1.10 - Predicted Probabilities of Poor Women with High and Low Schooling answering "good" or "very good health" per Region and Race

	Prob(SAH>=4)	**North**	**Northeast**	**Midwest**	**South**	**Southeast**	**Diff.Sth-North**
white	High schooling	0.7195	0.7479	0.758	0.7545	0.781	6.15%
	Low schooling	0.649	0.6805	0.6918	0.688	0.718	6.9%
non-white	High schooling	0.6968	0.7264	0.7369	0.7332	0.761	6.42%
	Low schooling	0.6243	0.6566	0.6683	0.6642	0.6954	7.11%

Souce: PNAD 2008

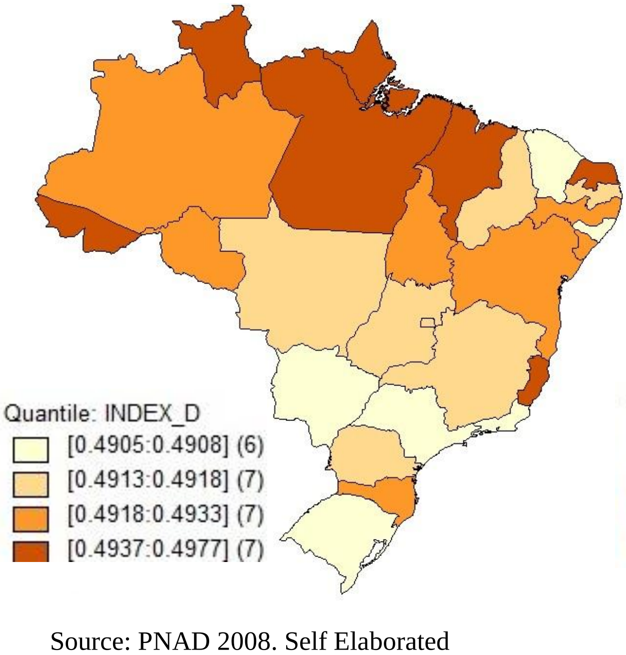
Regarding the propensity of saying good health, we can notice again, the disparity among Brazilian regions. Such gap is even wider for the most disadvantaged profiles (as showed in the last column of Table 1.10) if we compare North and Southeast regions. In addition, it noteworthy that, the white women hold greater probability of responding better health than non-white, if we compare those with the same schooling status. However, the non white women with high schooling, present better perceived health than the white women with low schooling, regardless the region.

Finally, a persistent gap between schooling groups – mainly in North and Northeast – is observed, no matter the race. Such evidence possibly suggests the important role of education over women´s health in the sense of "compensating" the lack of financial resources. Also, education is strongly linked with health literacy, which means, amongst other things, a more effective access to health information and a capability to utilize the health services more efficiently (Nutbean, 1998 APUD Social Determinants Of Health: SACOSS Information Paper, 2002).

Lastly, we report, though a map illustration, the Index-D (Figure 1.16), and the difference in probability of responding “good” or “very good” health (Figure 1.17), between the poorest quintile (quintile 1) and the richest one (quintile 5). We consider only women aging 16-65 years old, per state.

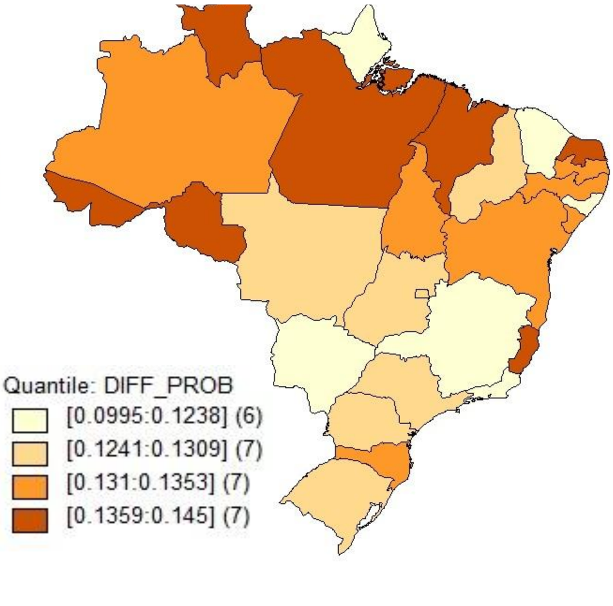
The darker areas represent respectively higher Index-D (Figure 1.16) and greater difference in the propensity of answering ‘good’ or ‘very good’ between the poorest and richest women population (Figure 1.17). We can observe that the North and Northeast regions hold, on average, states with higher inequalities than the South and Southeast ones; the same pattern is verified through Figure 1.17, i.e., the poorest women in most North´s and Northeast´s states perceive their health at a fairly lower level than the better-off group; whereas the same poorest women population from Southeast and South regions despite of answering poorer health than the richest ones, the difference is not too prominent.

Figure 1.16 - Index-D per Brazilian States between the worst and better-off income groups - 2008



Source: PNAD 2008. Self Elaborated

Figure 1.17 - Difference in predicted probabilities of answering “good” or “very good health” between the worst and better-off income groups per Brazilian States - 2008



Source: PNAD 2008. Self Elaborated

1.5 CONCLUSIONS

Given the increasingly interest of better understand the linkage between socioeconomic attributes of individuals and health status, several studies have been proposed a range of different health measures and alternative methodologies in order to assess socioeconomic health inequalities. The most used measure is the Self-Assessed Health, which provides a summary indicator of the respondent's own perception of her general health condition and has been shown a convenient and robust predictor of various health outcomes.

Once being of ordinal nature, a large body of work has attempted to transform this SAH variable into a cardinal one, since a continuous or dichotomous health indicator is required to estimate inequality indices and concentration curves. However, despite their property to summarize the distribution of SAH into a unique number, the suitability of imposing cardinality on what essentially is a categorical variable has been recently considered questionable.

Given this context, we proposed a novel approach – Index-D – that keeps the original ordinal nature of the SAH, without submit it to the cardinalization procedure. Our idea relied on generate SAH-predicted probabilities by means of an ordinal model estimation, as outcomes to be employed at a Gini measurement fitted for discrete probability functions. Since it is based on estimated probabilities, the Index-D enables us to assess variations in Self-Assessed Health for as many socioeconomic indicators as included in the probit model.

To illustrate its usefulness, we used data from PNAD database for the years 1998, 2003, and 2008. Besides income, we considered a range of socioeconomic and demographic attributes that has been widely reported as relevant determinants of Self Assessed Health.

Focusing on women reported health, the results suggested firstly, that there was a considerable decrease in the health inequality index from 1998 to 2008, for every profile investigated. As expected, the within index showed that income played an important influence to evaluate the own health. The lower the income quintile, the higher the assessed health inequality for women population.

Regarding (perceived) health disparities among regions, Southeast presented, in most analysis, the lowest Index-D, whereas the North region obtained the highest scores, no matter the profile under consideration. Furthermore, the predicted probabilities of answering better health clearly pointed out for a pro-rich health inequalities in Brazil, since the most disadvantage socioeconomic situations analyzed, were associated with greater probabilities of perceiving poorer health, irrespectively of the year and the region considered.

Along these lines, the method proposed here holds the attractive feature of exploring the pertinent health information provided by the SAH indicator, without incurring in the limitations of the cardinality assumption; and avoiding the inappropriateness of using external scores. In addition, the results found here aggregate empirical health information still scarcely investigated in the country. Finally, the simple and straight application of Index-D offers a potential tool to explore not only health inequalities, but also, any inequality analysis based on ordered variables.

2. PEOPLE AND WHERE THEY LIVE: THE ROLE OF MUNICIPALITIES ON SELF REPORTED HEALTH IN BRAZIL

2.1 INTRODUCTION

According to the World Health Organization (WHO, 1978), health can be defined not only in terms of absence of disease, injury or infirmity, but also, as a state of mental, physical and social well-being. Over the last decades, many studies have emphasized the role of social circumstances on health status. The tight link between health and a wide range of socioeconomic, environmental and demographics factors have been increasingly recognized and profferan alternative perspective on how to consider public health, social justice and even restructuring of the health care system (Daniels et. al., 2004). The increasingly acknowledgement that health is also a result of cumulative experience of social conditions and exposure to environmental contexts throughout the life course, has been leading to a ‘renewed interest’ (Anand and Peter, 2004) and a growing concern for Global Organizations (World Bank and OECD[1]).

The socio-economic determinants of health have been researched extensively, and health inequalities arises as a remarkable implication, since there are consistent evidences indicating that people from less favourable socioeconomic groups are more likely to suffer from higher rates of illness and mortality than the better off (see Kaplan, 1996; Wagstaff, 2000; World Bank, 1997). Individual and household poverty has been consistently shown as a risk factor for asthma and respiratory infections (WHO, 2012), coronary heart disease (Hart et.al.,1997), diabetes (Risteet. al. 2001) and homicide (Singh et al.; 2012) Therefore, it is acknowledged and well established by literature, a set of variables that most affect and influence the population’s health and wellbeing.

The traditional socioeconomic characteristics, in conjunction with the demographic ones, such as income, education, work status, access to information and appropriate health care, living and working conditions, age, gender and race, compose the set of factors most recurrent on literature. However, the set of potential socioeconomic determinants on health status lies beyond individual level effect. In recent years, an increasing body of studies has highlighted the important role of macro-level variables to understand social health inequality along with

[1] Organization for Economic Co-operation and Development

individual ones. Measurements reflecting deprivation and poverty at area level have reported an interconnectedness with health outcomes over and above individual attributes (Diez-Roux,1998; Macintyre et.al. 1993; Kaplan, 1996).

The geographic-level aspects involve a range of dimensions, ranging from physical characteristics of the area – such as location and climate (Bloom and Sachs 1998, APUD Wagstaff, 2001), to the infrastructure offered (Macintyre et. al., 2002) such as health services (quantity and quality), sanitation, water supply, roads, and so forth. One interesting hypothesis is that the presence of favorable aspects, like low crime rates, street cleanness and lightening, recreation places, among other amenities in the region under analysis mitigate the effects of unfavorable individual circumstances over health; whereas the adverse ones, such as pollution, lack of sanitation or low accessibility to urban facilities, amplify the already perverse influence of deprived individual characteristics on health status (Macintyre et. al., 2002; Kennedy et al., 1998). In California, for instance, Haan, Kaplan and Camacho (1987) found that people from poor areas experienced higher mortality rates (after controlling proper age, race and sex) than the population from non-poverty areas. Such risk of death persisted even after socioeconomic and behavioral adjustments. Similar findings have been reported by Humphries and Carr-Hill (1991); Jones and Duncan (1995) and Duncan et.al. (1993), supporting the hypothesis of the social environment´s influence over health, independently of the individual-level.

There is an important concern to take into account when it comes to analyze health status in space: the distinction between compositional and contextual effects (Curtis and Jones, 1998; Duncan et. al., 1993). The compositional effect means that the observed variation in health status is only due to population level characteristics, regardless where they live (Curtis and Jones, 1998). The analysis of context effect, on the other hand, aims to emphasize that the health status of an individual partially mirrors the influence of elements belonging the region of residence. In other words, this effect may represent the variability in health among people who hold similar socioeconomic attributes but belong to different regions; or alternatively, no matter the personal attributes, the population is affected by the exposure to environmental variables.

There is an intense debate on international literature about the adequate spatial scale for health analysis. It is often argued that contextual associations with health are most investigated using data referent to small areas (such as neighborhoods and communities); mainly in virtue of data availability (Diez Roux, 2008)and homogeneity ofsocioeconomic information (Curtis and Jones, 1998).As a result, there is an easily-noted prevalence of administrative areas (the Britain´s Electoral Wards) and neighborhoods, to investigate health outcomes through multilevel models. However, arguments in favor of analyzing larger areas also exist. As pointed out by Diez Roux (2008), neighborhoods "may not be the most relevant contexts for many health outcomes", i.e., there is also the necessity to research other "policy-relevant unities".

Following this reasoning, Curtis and Jones (1998) argue that effects of environmental factors, such as water and air quality, climate and degrees of urbanization, may not be well identified utilizing small area information, since they operate over wider area levels. Examples rely on studies that found significant area-level evidences by comparing rural-urban areas (Phillimore and Reading, 1992); north-south regions (Sloggett and Joshis, 1994); county districts (Wiggins et.al.,2002) and Federal States (Kennedy et. al.;1998). In fact, the most plausible hypothesis is that the contextual effects probably reflect an outcome of more than one geographic scale.

In order to take into account the importance of analyzing the simultaneous effects of both individual and contextual attributes over perceived health, the objective of this work relies on identifying in what extent the variance on Self Assessed Health in Brazil is a result of the context where the people live. Our geographic units will correspond to the representative sample of municipalities[2] available on the Brazilian National Household Surveys (PNAD) of the year 2008; along with a range of variables at individual level, including the Self Assessed Health.

Once the analysis will be extended to the whole country, the socioeconomic and health differences between municipalities may be further explained by the rich setting of political-institutional and cultural differences among the Federal States. In this context, we extent our final analysis to a third level, allowing the SAH to vary also between the States.

[2]As we will follow explain, we take into account only municipalities from metropolitan regions and that ones considered "self-representatives" by PNAD survey.

Being a country filled of social inequalities that extend beyond the individual dimension, we assume that the urban municipalities that we will take into account, represent a rich source of ecological exposure analysis, once they may reflect relevant disparities in a set of socioeconomic dimensions, as infra structure, urban facilities, social opportunities frameworks, etc. Moreover, this study aggregates evidences scarcely researched in Brazil. Not to mention the great policy relevance, given the importance of health status in a country where 74% of the population[3] (which is in aging process) uses public health services and where there are still statistics of infant mortality due to avoidable causes, such as nutrition deficiency, diarrhea and infections (Neri, 2007).

The following sections are organized as follows: in the next section we review the international literature; in the third section, we describe the dataset; the fourth section brings the Multilevel Model approach; the main results are presented and discussed in the fifth section and in the last one we present our final remarks.

3According to the Brazilian National Sample Household Survey (2008)

2.2 LITERATURE REVIEW

A large body of evidence from a wide variety of settings has documented what has come to be known as the *socioeconomic determinants of health*. The theory model that underlies such evidences relies on the causal linkages of demographic characteristics and socioeconomic circumstances over the occurrence of physical and emotional health problems on people[4]. An issue that has been drawn renewed interest and attention of much researchers and specialists of the field since early 1990s, is whether the social characteristics of the area where the individual live, exert influence on health and well-being as well as personal circumstances (Macintyre et.al. 2002). Indeed, research on socio-economic health inequalities has often invoked exposure to physical, social and biological risks as part of their explanatory theory (Curtis and Jones, 1998).

Hence, there is a relevant theoretical framework developed by this literature which would seem to support the reasoning that the contextual effects related to the place where the people are inserted in, have a relevant role over health variation. We first introduce, in a general basis, a conceptual support provided by Macintyre et. al., (1993) and Macintyre et. al (2002). The authors classify a set of physical and socio-environmental factors that might affect health outcomes into five broad dimensions: i.*Physical features of the environment*[5]; such as the quality of water and air, latitude and climate; which, in turn, are shared by the entire population of a wide area, like cities; ii. *The availability of healthy/unhealthy environments at home, at work and at play*, which is more related to secure/non secure of housing, place of work, and healthy recreation; iii.*The existence of public services*, such as health services, education, transport, street cleanness and lightening, security, etc.; iv. *Socio-Cultural features*, which include the political, economic, and ethnical circumstances that were shaped over time in a determined place; degree of community integration[6]; levels of crime and other personal

4See WHO (2012) and SACOSS (2008) for the discussion about the pathways of causal relation between health and socioeconomic status.

5 This dimension is also addressed by Curtis and Jones (1998), with the so-called 'spatial patterning and diffusion of physical and biological risk factors'.

6 This feature has been been subject of more recent studies. The well known "social capital" classifies itself in a more sociological approach, and has been considered an "inherently ecological" feature of places of residence. In this sense, studies as Kim and Kawachi, 2006; Habibov et. al., 2011; Han et.al.2012 highlights the mechanism by which social capital may reflect positively over health. Kakwani, 2006 for example, describes some potential mechanisms behind this relation:"diffusion of information about health promotion, maintenance of healthy behavioral norms through informal social control, promotion of access to local services and amenities, and psychosocial processes that provide affective support and mutual respect".

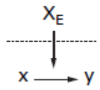
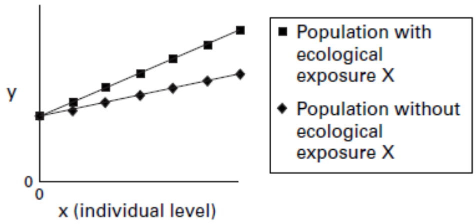
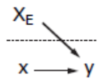
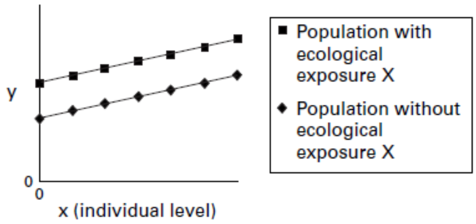
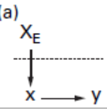
safety threats; v. The *reputation of a neighborhood*, or how the residents perceive the area they live, once it may affect the self-esteem and morale of the residents.

As reported by WHO (2012), not only people's exposure but also perception of the physical and social environmental hazard of where they reside may impact on their health, and even longevity. The 'psychosocial stress' has been interpreted more recently as an important determinant of health vulnerability against environmental risks. As argued by Duncan, Jones and Moon (1998), the context where a person lives can "shape and structure a set of opportunities and constraints of daily living, which exerts relevant effect over psychological wellbeing". A study developed by Rauh et. al (2008) APUD WHO (2012) for example, shows that overcrowding households, improper litter disposal or local busy transportation routes may cause chronic stress. This psychological dimension of environmental exposure can, still, be associated with the functioning of immune and inflammatory system response, leading to a higher vulnerability within unhealthy places (WHO, 2012).

All these range of environmental factors may exert influence on mental and physical health through different causal pathways. Moreover, the interaction with individual-level characteristics may lead to a greater (lower) health impact depending on the level of disadvantage (advantage) that the individual presents. The explanation relies on the higher (lower) individual vulnerability against environmental stresses and hazards: the so-called "exposure–response function" (WHO, 2012). As a matter of fact, socioeconomic disadvantaged groups are more often exposed to environmental health threats than the better off, possibly resulting in feelings of insecurity and inferiority and inducing unhealthy behavior and 'stress-related biological reactions' (Elstad, 2011). As a result, socio-demographic inequalities in health can also arise from different environmental risks exposure (WHO, 2012).

Regarding this issue, Blakely and Alistair (2000) provides three main causal pathways by which ecological exposures "X" may affect health outcome "y", taking into account the interactive role of individual level attributes "x": i. direct cross level effect; ii. cross level effect modification; iii. indirect cross level effect. The illustration is presented in the following Figure 2.1.

Figure 2.1: "*Three types of ecological effects*"

Ecological effect	*Example*	*Possible graphical representation*
1 Cross level effect modification 	*Example:* Income inequality (X_E) modifies the effect of social class (x) on the individual health (y).	
2 Direct cross level effect 	*Example:* Income inequality (X_E) directly affects individual health (y).	
3 Indirect cross level effect (a) 	*Example a:* Community tobacco control policies (X_E) may affect individual smoking (x), which in turn affects individual health (y).	Regression lines for individual level outcome (y) on individual level exposure (x) will not vary by population with varying ecological exposures (X), as any apparent effect of the ecological exposure(s) is explained away by including the relevant individual level exposures (x).

Source: Blakely T. and Alistair J. (2000)

Besides the possible direct effect, the reasoning that the area where a person lives may be on the causal pathway between individual´s social condition and health is also identified by Blakely and Alistair (2000). The "cross level effect modification" means that environmental circumstances may interact with individual level circumstances intensifying or reducing the negative (or positive) influence of personal characteristics on health outcomes, as suggested by WHO (2012). This effect is subjected to under estimation in quantitative research (as well as indirect effect), since it "ultimately operates at individual level" (Diex-Roux, 1998), i.e., individual variables once included in regression might "steal" such an effect from area level analysis. The "indirect cross level effect" of "X" on "y", on the other hand, relies on the conception that the area-level can affect, among other personal features, behavior patterns (represented by "x" in Figure 2.1) – such as alcohol consumption, diet or physical exercise practice – which, in turn, play a direct influence on health outcomes. As emphasized by Barufi et.al. (2012), geographical and cultural features can differ hugely between municipalities in Brazil, and may have a relevant influence over health habits.

A large amount of empirical evidences has reported significant geographical variations in a wide range of health outcomes, as for instance, in coronary disease mortality (Diez Roux et.al,1997); morbidity (Jones and Ducan,1995); depression (Yen and Kaplan, 1999); and behavior patterns, such as alcohol consumption (Ecob and Macintyre, 2000 APUD Macintyre et. al., 2002), diet and physical activities (Cubbin et. al., 2006; Karvonen and Rimpela, 1996, APUD Ellaway and Macyntire, 2009) and self reported health (Jones and Duncan, 1995; Humphreysand Hill, 1991).

A classic example relies on Blaxter (1990). Using a Britain database, the author explored different dimension of health (illness, fitness and psychosocial health) among manual and non manual class groups living in distinct areas. The author pointed out that cities seem to affect negatively the psychological and illness dimensions over non manual workers; whereas 'industrial areas' play damaging effect over health of manual workers. Macintyre et. al. (1993), through a qualitative comparison between two socioeconomic distinct areas in Glasgow, found that no matter the personal characteristics, the 'opportunity framework' in the poorer area is 'less conducive' to better health than in the better-off area. Additionally, the authors argue that thearea deprivation is 'magnified' once combined with individual disadvantage conditions.

The hypothesis that health outcomes are influenced by macro-level factors as well as by individual attributes has been tested by a large number of studies and a fundamental concern arises from this literature: the relevance to take into account, simultaneously, the compositional and contextual effects upon health analysis. Duncan, Jones and Moon (1993) reported an intuitive example about this issue. They argued that places with high levels of smoking may represent areas merely composed by people with a predisposition to smoke. On the other hand, the place may be 'impregnated'– regardless the individual characteristics – by a contextual, or even historical, factor, such as "a regional culture that encourages smoking". In this sense, a large number of studies (e.g. Jones and Duncan, 1995; Humphreys and Hill, 1991; Stafford et. al. 2001) employ and recommend multi-level modeling, which allows the researchers to quantify the 'size' of the area effect separately to the individual influence.

Most of these studies report that individual health variance is explained mostly by the first level (individual or compositional effect) and that, in spite of small, the second level (area effect) presents a persistent significant variation, even after controlling individual variables. As argued by Blakely and Alistair (2000), such a small ecological effect may be, indeed, 'the tip of the iceberg' and should not be ignored or considered irrelevant.

Stafford et. al.(2001) for instance, in an attempt to answer the area effect on self reported health, used data from Whitehall II (1991) through a two level binary model, where the second level were the Wards[7] from London, England. They show that the self reported health varies significantly across the wards, presenting 20% of variance amongst men in the null model and 18% after individual variables control, representing for the authors, the 'tendency for similar individuals to live in the same areas'. Nevertheless, great part of the variation on the second level remained. Thus, the authors, further analyzing such residual variation, found that, after individual socioeconomic control, those living in most deprived areas were 1.29 times as likely to assess poor heath than those who live in advantaged areas, and that those who reported serious problems with the neighborhood were 2.73 more prone to respond poorer health than people who didn´t complain about their neighborhood.

Similar results were found by Jones and Duncan (1995) through three different measures of chronic illness from Health and Lifestyle Survey, British, for the years1984/1985. They computed that 7% of the overall residual variance lies between the wards, irrespective of individuals characteristics. In an attempt to explain such residual, the authors have taken into account indicators of deprivation and income inequality at area level. The results showed that, in general, people from places with low average income and high deprivation[8] suffer more health damage, regardless the health measure applied.

On the other hand, studies as Sloggett & Joshi, (1994) and Davey Smith et al., (1995) APUD Macintyre et.al., (2002), reported that after a range of individual characteristics control, the area level variance lose its statistical significance. The former authors argue that area deprivation appears to be adequately assessed by personal or household circumstances, which are themselves associated with income. About such conflicting evidences, Macintyre et.al. (2002) points out that the distinction between composition and context effects "may not be as conceptually clear or as useful as may appear at first glance". Indeed, as Diex-Roux (1998) argue, individual attributes in multilevel models may be 'intervening' variables rather than 'confounders' on the pathway between place and health – the "*cross level effect modification*", briefly discussed above.

[7]Wards in Britain are an electoral district within a municipality.
[8]"Place-based deprivation refers to poor access to specific goods and services" (Jones and Duncan, 1995).

In Brazil, few studies tried to analyze the area effect over health outcomes. Noronha and Andrade (2005) for example, using multilevel logit estimation and PNAD 1998, found that individuals who live in places with lower income inequalities have more chances of assessing better health. Soares (2006) in turn, analyzed the reductions in mortality between 1970 and 2000 across Brazilian municipalities. The author shows throughout a dynamic panel setting an interesting result: 71% of the within municipality variation in mortality is explained by 33% of changes in income per capita, 16% in access to water, 16% in illiteracy and 6% in sanitation. It means that beyond income, the changes in life expectancy in Brazil are consistently explained by urban infrastructure and education. Results found by Noronha and Andrade (2002) suggest that health inequality[9] differ among federal states. Distrito Federal had the highest inequality, whereas Rio de Janeiro and São Paulo presented the lowest ones.

2.2.1 INTRODUCING THE HEALTH OUTCOME AND THE EXPLANATORY VARIABLES: A BRIEF DISCUSSION BASED ON LITERATURE

Before present the descriptive statistic, it seems appropriate to briefly introduce our health variable, along with the contextual ones, based on what has been discussed by the researchers of the field.

2.2.1.1 The Outcome Variable: Self Assessed Health

Recorded on an ordinal scale (i.e. “poor”, “fair”, “good”, “very good”, and “excellent”), SAH provides a summary measure of the respondent’s own perception of his/her general health condition. It integrates not only an indication of the burden of disease experienced by the person, but also it embodies emotional and social aspects of health and well-being, since its assessment is subjected to contextual effects, as socioeconomic and cultural elements experienced by the person. As a result, it is recommended to social analysis (Tubeuf, 2008; Curtis and Jones, 1998). Due its facility to be collected, it is the most common measurement of health in household surveys and has been indicated as a robust predictor of objective health outcomes, as subsequent illness and premature mortality (Idler and Benyamini, 1997; van Doorslaer).

[9] The health inequality took into account different health indicators: Mobility, Chronic disease and Self assessed health

In spite of holding a set of advantages, SAH presents challenges in terms of empirical analysis. Being subjective, its assessment is based on different reference points (Groot, 2000). The so-called reporting bias implies that individuals belonging different socioeconomic and demographic groups may evaluate their own health differently even if their "true" health is similar. Nevertheless, in accordance with Nogueira (2008), we assume that the SAH´s subjectivity is not a limitation, but rather, a favorable feature for this sort of analysis. As just commented above, illness-health perceptions and experiences lived by the person are also a function of individuals and contextual circumstances, consequently better captured if the nature of the variable is indeed, more sensible to these social dimensions.

2.1.1.2 Contextual Variables

Many researchers have argued that the status of well being of individuals is not only affected by their personal attributes but is also a result of the socioeconomic characteristics of the groups to which they belong (Pickett and Pearl, 2001). Diez-Roux (1998), for instance, emphasizes that 'group-level'[10]variables in hierarchical analysis detain relevant information not provided by individual data. They argue that the mean income of a neighborhood may be an indicator of the existence (or lack) of potential factors and circumstances related to health, such as recreation facilities, school quality and environmental conditions. Put in another way, the health of individuals holding similar socioeconomic characteristics may vary according to the 'socioeconomic setting' in which they live (Curtis and Jones, 1998).

Along these lines, we extent such an argument to the municipality scale, in the sense that higher socioeconomic development, reflected by indicators as mean income[11] and schooling, may imply for example, better public services. We explore a set of contextual variables at municipal level that may indicate, the degree of infra structure and collective social environment of the region. In this sense, we include in our second level analysis the proportion of individuals over 25 years old who attended less than five years of school; the ratio of houses holding piped water and sewage[12]per municipality; the Gini coefficient of each

[10]Several different terms have been used as synonyms for group-level variables in these types of analyses, including ecological variables, macro-level, contextual and aggregate variables (Diez-Roux, 1998)
[11] Due high correlation between mean income and mean schooling, we decided to keep only the second variable among the municipal-level covariates.
[12] Defined by sewage connected to the main sewer or connected to fosse.

municipality, which is particularly intended to reflect the well documented (Kennedy et. al., 1998) influence of the income inequality on health variance; and finally we try to verify the potential role of "Programa saúde da Família[13]", characterized by the proportion of families who are assisted by the program.

It should be pointed out again, that the presence of heterogeneity amongst individuals within a geographic unit, as a municipality, may be too complex to be represented in "mean" terms. Also, some group-level variables used here, present significant correlation (see correlation matrix attached), since deprived (developed) socioeconomic indicators tend to be concentrated into the same areas.

[13]"*Saúde da Família*" is understood as a strategy for reorienting the assistential model, operationalized through the implementation of multidisciplinary teams in primary health care units. These teams are responsible for monitoring a defined number of families, located in a delimited geographical area. The teams work with actions of health promotion, prevention, recovery, rehabilitation of diseases and more frequent disorders, and in maintaining the health of this community (http://portal.saude.gov.br/portal/saude/cidadao/area.cfm?id_area=149)

2.3 DATA AND METHOD

2.3.1 DATA

The dataset employed in our analysis relies on the Brazilian National Household Survey (Pesquisa Nacional por Amostra de Domicílio-PNAD) referent to the year 2008.The PNAD database is an annual[14] national survey of households samples that covers all urban and most of rural areas[15] of Brazil. It collects an extensive set of demographic and socioeconomic variables of the population, along with information related to the household conditions–which includes basic infrastructure, such as piped water and sewage. PNAD database also holds an additional questionnaire addressing specific social subjects that vary periodically according with the requirements of the country. Particularly for the years 1998, 2003 and 2008, these supplements were based on a range of information related to the population´s health, which includes our key variable Self Assessed Health. Our final dataset totalizes198,461 individuals.

2.3.1.1 The Geographic Area: The Pnad´s Primary Sampling Units (UPAs)

We employ in the present study, the most disaggregated area available for statistical analysis regarding Self Assessed Health in Brazil: The PNAD´s "primary sampling units". Each primary sampling unity (UPA) corresponds to one municipality within the PNAD´s sample[16]. It is adopted the same territorial division utilized by the Demographic Census of IBGE. They are selected at the beginning of every decade and are followed along this referred period. The sample of municipalities is chosen to be representative for the whole country (Foguel and Barros, 2010). Particularly for the current decade, the survey holds 817 municipalities that are classified into three groups: 139 belonging metropolitan areas, 134 considered by IBGE to be large in population terms[17] (the "self-representative" ones) and 544 representing the remaining smaller municipalities[18].

[14]Except for the years when the national survey (census) is collected and in 1994 for budget constraint
[15]The rural Amazon areas are excluded
[16] The UPAs classified as "self representatives".
[17] See the criterion employed by IBGE to select the self-representative municipalities in Souza and Silva (2003).
[18] The name of the municipalities is not divulgated by PNAD survey.

We excluded from our dataset, the 544 smaller UPAs, keeping only the self representatives´ and metropolitan regions´municipalities from the whole country[19]. This choice is supported by the idea of keeping only population from the most representative urban environments of the country. Also, we avoid the municipalities that hold a small number of individual observations. From now on, we will call our sample of UPAs as simply "municipalities". Table 2.1 show the selected municipalities distributed across the Federal States and their respective population sample:

Table 2.1 - Number of individuals and self representatives´ and metropolitan regions' municipalities per Federative State

states	**nº of Municipalities**	**sample**	***states***	**nº of Municipalities**	**sample**
Rondônia	5	3,657	*Espírito santo*	7	4,398
Pará	7	13,730	*Rio de Janeiro*	29	25,248
Tocantins	3	1,885	*São Paulo*	50	30,574
Ceará	14	17,738	*Paraná*	19	11,514
Rio G. Norte	3	2,762	*Santa Catarina*	8	3,515
Pernambuco	15	18,519	*Rio G. sul*	33	21,860
Sergipe	3	2,800	*Mato G. do Sul*	4	4,083
Bahia	17	20,411	*Mato Grosso*	5	3,286
Minas Gerais	25	17,936	*Goiás*	13	10,060

Total

states	**Municipalities**	**Sample**
18	**260**	**213,976**

Source: PNAD 2008

2.3.1.2 Individual and Contextual Variables

Representing the fixed effect at the first level of our model, the individual's variables have the important role of capturing the compositional effects in an attempt to provide and interpret the contextual results more accurately. The first set of variables regards the demographic characteristics. Besides the important age control, we also include race, categorized by 'one', respondents who declared themselves of being white or yellow, and 'zero', brown or black. The income measure relies on a continuous variable representing the household per capita income in natural logarithm scale; the schooling attainment is

[19]The states of Acre, Amazonas, Roraima, Amapá, Maranhão, Piauí, Alagoas and Distrito Federal were extracted from our sample since they only have one or two self-representative and/or metropolitan regions´ municipalities within them.This probably biases our approach as long as there is no municipality variance within the states.

represented by the number of years that the individual attended school, ranging from no education to more than 16 years.

It has been also shown, that health may be greatly affected by work status (Stafford, 2001; Blaxter, 1990). We consider, thus, a binary variable that identifies if the person worked in thelast week of reference, and it corresponds to a proxy for being currently employed. A dummy, representing those who hold private health insurance – and, on the contrary, zero for those who depend on public services[20] – is also a relevant variables to account for, especially in Brazil, where both the quantity and quality of the service offered by the public health system varies hugely across country. Finally, as highlighted by (WHO, 2012), overcrowded households along with other housing conditions variables, has been associated with respiratory problems. In this sense, we add a last explicative variable representing the number of residents in the households. The descriptive statistic of individuals characteristics, along with the contextual variables, already presented in the previews section, are reported in Table 2.2.

Table 2.2 - Descriptive Statistics of individuals and contextual variables

Individual Variables	***code***	**Mean**	**SD**	**Min**	**Max**	**n° of obs**
Gender (male=1)	*gender*	0.477	0.499	0	1	198461
Age	*age*	31.916	20.170	0	108	198461
Race (white=1)	*race*	0.491	0.500	0	1	197650
household *per capita* income	*ln_pcincome*	6.172	1.132	0.9163	11.9184	208051
Years of schooling	*schooling*	7.797	4.745	1	16	197754
Worked in the week of reference (yes=1)	*employed*	0.545	0.498	0	1	169512
Own private health insurance (yes=1)	*health_ins*	0.337	0.473	0	1	198461
Total household´s residents	*nº resid*	3.990	1.782	1	20	198345
Contextual Variables	***code***	**Mean**	**SD**	**Min**	**Max**	**n° of obs**
% sewage [1]	*%water*	0.819	0.199	0	1	213976
% piped water	*%sewage*	0.970	0.050	0.4099	1	213976
mean income	*mn_income*	6.171	0.317	5.229	7.36014	213976
% of people over 25 years old with low schooling[2]	*lowschooling*	0.247	0.074	0.03306	0.666	213976
% Health Program	*%familyprog*	0.391	0.248	0	0.9555	213976
Gini Coefficient	*Gini*	0.544	0.003	0.5425	0.6104	213976

Notes: 1.There is one municipality in Goiás, where the entire sample (255 respondents) declared no sewage in their household; 2. People with five or less years of schooling.
Source: PNAD 2008

[20] We assume that those who answered do not hold private health plan in general use SUS.

For the most representative municipalities of the country, the average profile of the individuals sample follows the expected. A minority holds private health care; a high variance in the income per capita; the gender and race well balanced; and more than half of the respondents worked in the week of reference. Regarding the contextual indicators, there is a considerable variance in most of the aggregated variables at municipal level. The rate of sewage for example, ranges from 0 to 100% of coverage, while the proportion of adults over 25 years old with low schooling varies from 3% to more than half of the adult population presenting low schooling[21].

2.3.1.3 Self Assessed Health

Given the previews introduction about our health outcome, the following Table 2.3 provides the frequency of SAH responses, along with its mean, for both entire and detailed sample, according to individual´s socioeconomic and demographic profile and contextual variables.

Table 2.3: Descriptive Analysis of Self Assessed Health according to Individuals Socioeconomic Characteristics and Contextual Variables

Self Assessed Health		***Very Poor***	***Poor***	***Fair***	***Good***	***Very Good***	***Total***	**Mean**	**St.Dev**
Total Sample	**Freq.**	1,420	5,110	35,585	107,952	48,397	198,464	**3.991**	**0.768**
	%	0.72	2.57	17.93	54.39	24.39	100.00		
				Individual Variables (%)					
Gender		***Very Poor***	***Poor***	***Fair***	***Good***	***Very Good***		**Mean**	**St.Dev.**
Men		0.62	2.37	15.9	55.05	26.07		4.035	0.754
Women		0.8	2.76	19.79	53.8	22.85		3.951	0.778
Age Group									
<15		0.09	0.49	7.76	55.58	36.09		4.2707	0.626
>15 e <30		0.21	0.91	10.78	59.68	28.43		4.151	0.6515
>30 e <45		0.53	1.97	18.92	57.75	20.83		3.963	0.723
>45 e <60		1.37	5.11	30.02	48.7	14.8		3.704	0.8298
>60		2.84	9.24	39.8	38.9	9.22		3.4241	0.885
Race									
White		0.68	2.33	15.69	53.02	28.28		4.059	0.7688
non-white		0.73	2.78	20	55.79	20.69		3.929	0.76
Work status									
trab (16-65)		0.33	1.59	17.2	57.7	23.17		4.017	0.704
non-trab		1.42	4.61	23.63	51.55	18.8		3.817	0.838

[21] It is important to highlight the advertence given by Foguel and Barros (2008) that the mean sample values at the municipal level may be subjected to sampling error, generating possible measurement error problems.

health insurance							
have	0.47	1.56	14.03	53.12	30.82	4.122	0.7335
don´t have	0.84	3.09	19.92	55.04	21.11	3.924	0.777
Income							
1-quintile	0.72	3.05	19.47	57.04	19.72	3.919	0.7564
5-quintile	0.58	1.66	13.89	51.79	32.07	4.131	0.749
Schooling							
1-quintile	1.17	3.92	18.51	49.89	26.5	3.96	0.8434
5-quintile	0.25	0.92	9.94	54.1	34.78	4.222	0.6765
			Contextual Variables (%)				
% pipedwater	***Very Poor***	***Poor***	***Fair***	***Good***	***Very Good***	**Mean**	**St.Dev.**
1 quintile	0.68	3.65	21.12	57.6	16.95	3.864	0.7549
5 quintile	0.66	2.43	16.24	52.52	28.16	4.05	0.773
% sewage							
1 quintile	0.55	3.02	19.74	54.94	21.75	3.943	0.7627
5 quintile	0.76	2.27	16.45	52.98	27.54	4.042	0.773
Families´ HealthProg							
1 quintile	0.7	2.42	17.22	53.93	25.72	4.015	0.7678
5 quintile	0.66	2.87	18.84	54.95	22.68	3.961	0.766
mean income							
1 quintile	0.65	3.49	22.39	56.55	16.93	3.856	0.7551
5 quintile	0.65	2.07	15.11	51.24	30.93	4.097	0.7695
%lowschool.							
1 quintile	0.66	2.28	16.67	53.49	26.9	4.0369	0.7646
5 quintile	0.63	4.47	21.04	49.25	24.6	3.927	0.8289
Gini coefficient							
1 quintile	0.7	2.26	17.25	55.56	24.23	4.0036	0.7537
5 quintile	0.63	2.91	18.78	54.47	23.2	3.967	0.7693

Source: PNAD 2008

Following the majority of the evidences containing SAH, most of the population sample reported good or very good health. However, if we look this distribution for distinct socioeconomic profiles, a considerable difference in perceived health can be noted. For example, those who are not employed (between 16 and 65 years old), perceive very poor health four times more than the employed ones. Regarding the contextual indicators, remarkable differences are also observed, as for instance, the distinction between the frequencies of ‘very good health’ responses from the first to the fifth sewage and income quintiles.

2.4 EMPIRICAL ESTRATEGY

Multilevel Modeling (Goldstein, 1986) or Hierarchical Regressions (Bryk and Raudenbush, 1987) are generally employed to analyze data that are shaped into a hierarchical structure, consisting of individuals clustered within macro units, which themselves may be nested within further higher level groups. As a result, a correlation structure arises, invalidating classical assumptions of independence and consequently OLS estimations (Rice and Jones, 1998). Moreover, the ‘richness’ of information potentially contained in hierarchical data is missed if only a single level model is considered (Rice and Jones, 1998).

In this sense, multilevel models technique provides an attractive approach to assess health variations without ignoring the contextual influence of where individuals are inserted in, and also, allows investigations of group-level characteristics that are potentially predictors of such differences. In addition to the multilevel structure, our dependent variable is recorded in an ordinal scale, leading us to the fact that regression analysis based on ordinal data – as well as other types of non-normal outcomes, as binary or rates – should bedesigned within a family of models so-called *generalized linear models* (O´Connell, 2010). Therefore, to fit appropriately the specific framework of our data, the generalized linear mixed model (GLMM) is strongly recommended (see details in Gibbons and Bock, 1987 and Rabe-Hesketh, 2005; for respectively modeling and estimation procedures regarding ordinal responses in multilevel contexts).

In this context, a *Two-Level Random-Intercept Ordered Probity Model*[22] is performed in the present paper using STATA 11 software package along with *GLLAMM* module[23] (Rabe-Hesketh 2005) for statistical computation[24].

Following Rabe-Hesketh (2003), the Two-Level Random-Intercept Ordered Probity Model is illustrated as follows:

[22]The last model of our analysis corresponds to a Three-Level Random-Intercept Model.
[23]For and information and clarifications about STATA commands, see “*GLLAMM companion* by Rabe-Hesketh, S. and Skrondal, A. (2012)
[24]To compute the parameters of a Random Component Probit Model, Rabe-Heskethet. al., (2002) demonstrated through simulations that adaptive quadrature is the most efficient method to numerically evaluate and maximize the marginal log likelihood. In this context, the present paper employs the *Gllamm* module that has been implemented by Rabe-Heskethet. al., (2002) and embrace the estimation of multilevel generalized linear mixed models using adaptive quadrature.

$$y_{ij}^{*} = \beta_{0j} + e_{ij}(1)$$

Where y_{ij}^{*} represents the latent health status of individuals i=1,2…n (level 1), clustered in municipalities j=1,2…m (level 2), that underlies the observed categorical SAH response:

$$SAH_{ij} = k \; if \; \lambda_{k-1} < y_{ij}^{*} < \lambda_{k},(2)$$

The λ's correspond to the estimable threshold values that mark the boundary between ordinal classes and are constant across items i. For instance, in the case of SAH, K=5,the levels are k=1="very poor", k=2="poor", k=3="fair", k=4="good", k=5="very good".The residual error term assumes normal distribution and variance equals 1, since it is based on a Probit Model ($e_{ij} \sim N(0,1)$).

Representing the random intercept or "group dependent intercept" (Crouchley et. al., 2009) at equation (1), β_{0j}, will vary between municipalities units. The so-called level-2 equation, in its simplest form, assumes the following function:

$$\beta_{0j} = \boldsymbol{\gamma}_{00} + u_{0j} \quad (3)$$

Where the intercept $\boldsymbol{\gamma}_{00}$ corresponds to the SAH's global average value and the error component u_j can be interpreted as the deviation of municipality j SAH mean from the grand mean. It is assumed to be independent of $e_{ij}$and normally distributed with mean 0 and unconditional variance σ_u^2: $(u_j \sim N(0, \sigma_u^2))$. The core of the analysis relies on observing if the variation of the self reported health at level-2 (σ_u^2) remains significant even if a range of relevant socioeconomic and demographic variables ($\boldsymbol{x}_{ij}$) are taken into account in the model, representing the compositional effect control of individuals i from a particular municipality j:

$$y_{ij}^{*} = \boldsymbol{\gamma}_{00} + \boldsymbol{\beta}_{1(1)}\boldsymbol{x}_{ij} + e_{ij} + u_{0j} \quad (4)$$

Where the variance of y_{ij}^{*}conditional on $\boldsymbol{x}_{ij}$ can be decomposed as the sum of level-1 and level-2 variances: $var(y_{ij}^{*}|\boldsymbol{x}_{ij}) = var(u_{0j}) + var(e_{ij}) = \sigma_u^2 + 1$ (5)

The most interesting result, and focus of this paper, is asking how much of the total residual variance can be attributed to level-2 units. The residual intra-class correlation coefficient (ICC) provides this answer:

$$\rho = Corr(y_{ij}^{*}|\boldsymbol{x}_{ij}) = \frac{\sigma_u^2}{\sigma_u^2 + 1} \quad (6)$$

Bringing to our context, ρ gives us the proportion of Self Reported Health variance that lies between the municipalities in Brazil. Still, to further examine any between-area health differences, we add to equation 4 the vector of contextual variables ($\boldsymbol{z}_j$)that might explain at least a fraction of the second-level variance[25]:

$$y_{ij}^{*} = \boldsymbol{\beta}_{1(1)}\boldsymbol{x}_{ij} + \boldsymbol{\gamma}_{1(2)}\boldsymbol{z}_j + e_{ij} + u_j (7)$$

Finally, one may argue the existence of a SAH pattern between the federal states, once the municipalities are clustered within them. In this sense, we further investigate the variance in SAH by estimating a fourth model with three random effect levels, which the third level corresponds to18 Federal States. The last model can be computed analogously to equation 3:

$$\beta_{js} = \boldsymbol{\gamma}_{0s} + u_{0js} \quad (8)$$

Where

$$\boldsymbol{\gamma}_{0s} = \boldsymbol{\tau}_{00} + \varepsilon_{00s} \quad (9)$$

Resulting in:

$$y_{ijs}^{*} = \boldsymbol{\tau}_{00} + \boldsymbol{\beta}_{1(1)}\boldsymbol{x}_{ij} + e_{ijs} + u_{0js} + \varepsilon_{00s} \quad (10)$$

Where the municipalities intercept is now varying between states, generating a new residual component (ε_{00s}) that presents in turn, a variance ofσ_{ε}^2 ($\varepsilon_{0s} \sim N(0, \sigma_{\varepsilon}^2)$. As a result, the total residual variance is decomposed into three parts:

$$var(y_{ijs}^{*}|\boldsymbol{x}_{ij}) = var(e_{ij}) + var(u_{0js}) + var(\varepsilon_{0s}) = 1 + \sigma_u^2 + \sigma_{\varepsilon}^2 \quad (11)$$

And the residual intraclass correlation coefficient of the states is estimated similarly to (6):

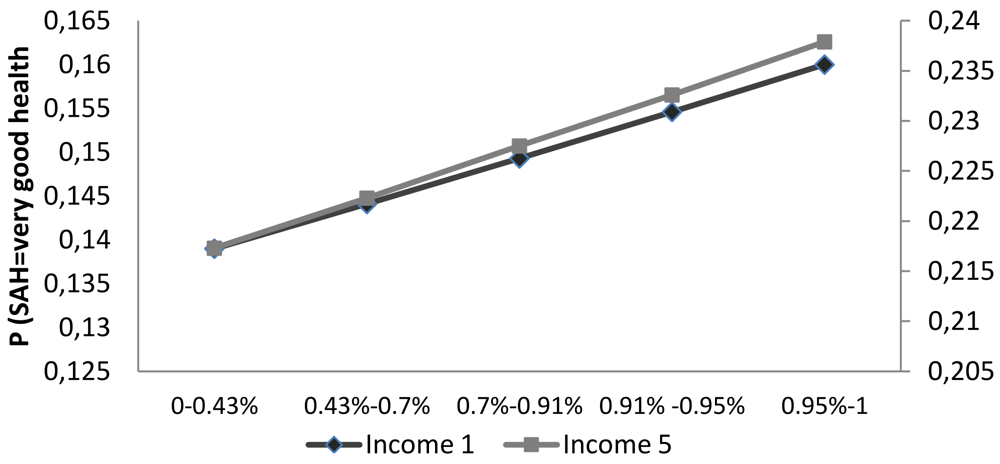
$$\gamma = Corr(y_{ijs}^{*}|\boldsymbol{x}_{ijs}) = \frac{\sigma_{\varepsilon}^2}{\sigma_u^2 + \sigma_{\varepsilon}^2 + 1} \quad (12)$$

[25] The intercept is omitted for identification purpose

2.5 RESULTS

To discuss our main results, we first proceed with a descriptive analysis usually employed by studies of the field, which relies on an interaction between individual social class position and area deprivation. In our case, we rank into quintiles the proportion of sewage coverage per municipality as a proxy[26] for the degree of infra structure and broad socioeconomic environment of the municipality. In crude terms, we assess in what extent and direction, the probability of individuals perceiving better health alters as long as we consider different socioeconomic municipalities´ groups. A simple Ordered Probit Model was estimated adjusted for individual level attributes[27]in order to predict the probability of saying "very good health". We analyze only the individuals from the worst (income 1) and better-off (income 5) income quintiles, as displayed in Figure 2.2:

Figure 2.2[1]- Probability of answering "very good health" of the worst off – income 1 – and better-off – income 5[2] – according to five groups of municipalities ranked into sewage coverage quintiles



Note: 1. the left vertical axis represents the predicted probability of the "income 1" group; whereas the right axis corresponds to the "income 5" group. 2.The first and fifth quintile of household per capita income
Source: PNAD 2008. Self elaborated

The results reported by Figure 2.2 reveal interesting evidence. As expected, there is a clear positive relation between answering better health and the upper "deprived" groups, regardless the individual class position. This same analysis was performed for other relevant socioeconomic aspects of an urban area: income inequality, proportion of adults with low

[26] Researchers (e.g. Duncan et. al., 1998) usually construct a "deprivation" index combining a set of indicators of the locality. In future investigations this procedure should be attempted.

[27]Age, race, gender, per capita income; years of schooling; work status; health insurance. Additionally, the mean income by municipality.

schooling, and piped water coverage[28].The increased propensity of saying better health, as we move from the less to the better-off municipalities´ group, persisted for all (see Figure 2.4 attached).

In rough terms, this suggests that municipalities with the lowest levels of sewage coverage have a population that on average present a diminished propensity of saying better health. This first evidence from our data suggests that the contextual level cannot be ignored, and introduce the following multilevel model analysis. Table 2.4 reports the main results.

Table 2.4 - Random-Intercept Ordered Probity Model: Fixed and Random effects of Self Assessed Health in Metropolitan and self-representatives municipalities of Brazil

Dependent Variable:Self Assessed Health	**Null Model**	**Model 1**		**Model 2**		**Model 3**	
Fixed Effect		Coeff	S.d	Coeff	S.d	Coeff	S.d
Individual (*Level-1*)							
Gender(male=1)	-	0.0838	0.0057	0.084	0.0057	0.0838	0.0057
Age	-	-0.0245	0.0002	-0.024	0.0002	-0.0245	0.0002
Race (white=1)	-	0.0954	0.0062	0.096	0.0062	0.0953	0.0062
ln_pcincome	-	0.1300	0.0036	0.130	0.0036	0.1297	0.0036
Schooling	-	0.0347	0.0008	0.035	0.0008	0.0348	0.0008
employed	-	0.0708	0.0060	0.071	0.0060	0.0709	0.0060
Health insurance	-	0.1212	0.0067	0.121	0.0067	0.1207	0.0067
nº of residents		-0.0031	0.0017	-0.003	0.0017	-0.0030	0.0017
Contextual Variables (*Level-2)*							
Gini	-	-	-	-0.3287	2.5701	-	-
lowschooling	-	-	-	**-0.3115**	**0.11502**	-	-
%Sewage	-	-	-	**0.268**	**0.0514**	-	-
%Water	-	-	-	0.131	0.1511	-	-
%family_program	-	-	-	**0.125**	**0.029**	-	-
Random Effect							
level-1 (Individuals)	1	1		1		1	
level-2 (Municipality)	0.1189*** (0.0054)	0.0769***	0.0041	0.0929***	0.0049	0.044***	0.0033
ρ	**11%**	**7%**		**9%**		**4%**	
level-3 (states)						0.1272***	0.0166
γ	-	-	-	-	-	**11%**	
log Likelihood	-218570	-166632.1		-166641.88		-166612.4	
n	198,464	163,664		163,664		163,664	

[28]We employ here the reasoning given by Pickett and Pearl (2001), i.e., we know that all these indicators are simple measures being used as proxies for complex concepts, that we intuitively understand but cannot measure accurately.

N°Municipalities	260	260	260	260

Note: The cut-off points were omitted
Source: PNAD 2008

Table 2.4 displays the multilevel estimation for the ordered variable SAH. The modeling procedure begins with a baseline model reporting only the random component. This strategy is recommended in order to give an idea about the upper bound of the contextual-level effect. As commented before, it is fairly plausible that individual information captures a fraction of the group-level effect, since the context where the people live may exerts an effect firstly over personal aspects – as behavioral patterns and habits – which ultimately influence health outcomes. As a result, they may be mediators rather than confounders (Diex Roux, 2008).

The intra class correlation (ρ) reported in the null model may suggest, roughly, that 11% of the Self Assessed Health variance lies between municipalities. However it is well knowledge that this are-level effect may be simply a result of individual attributes, since persons with similar socioeconomic characteristics tend to share the same areas. For this reason, Model 1 takes into account some important socioeconomic individual attributes in an attempt to separate out the compositional effect that probably contaminates our first result. In fact, the correlation coefficient decreases from 11% to 7%, which means that the individual socioeconomic circumstance is responsible for a reduction of 36% of the variance between municipalities found in the baseline model. Nevertheless, the remaining correlation of 7% is still statistical significant ($p<0.0001$). This result suggests, with other words, that there is a substantial systematic variance in SAH between municipalities which cannot be explained by the differing social composition of the areas, indicating that the perceived health of the individuals within a same city are somehow correlated in virtue of contextual features not observed in our data.

Regarding the individual covariates *per se*, we found the expected signals. Men are likely, on average, to respond better health than women; the older the person the higher the chances of saying poorer health; individuals who declared themselves as white or yellow has a slight greater probability of perceiving better health; the higher chances of responding good/very good health for those with better financial conditions, as well as for those with more years of schooling. The work status also seems to have an important role; along with holding health insurance, which may reflect both the people with better financial conditions (and can pay a private health insurance); and the quality distinction between private and public health services affecting how people evaluate their health. Lastly, we verify that number of residents within a household still present a slight negative correlation with SAH ($p < 0.10$).

In an attempt to explain at least a fraction of the contextual level variance, we further include a set of group-level covariates in Model 2. As already argued here, we assume that such aggregated variables, represent more than simply ‘summaries’ of individual attributes, and are used as an indicator of the socioeconomic environment of the region considered. The existence of a considerable proportion of adults older than 25 years old that attended the school only for less than six years (*lowschooling*)[29] seems to play a negative role on the reported health of the city, presenting a statistical significant coefficient ($p<0.005$). A possible explanation given by (Elstad, 2011) is that if there is a predominant high educational level in the surrounding area, the healthy behaviors might spread through the population more rapidly. Also, it can be related to the social capital, since the poorer and less educated populations, despite of being more at risk of health hazards, hold weak political power to contest the social vulnerability and demand public services (Curtis and Jones, 1998). We are aware that this reasoning is better fitted for smaller areas, but we cannot discard such hypothesis for a municipal scale.

The proportion of sewage coverage corresponds, as expected, to a positive contribution of the municipalities´ level over better health status. However, it is difficult to isolate its effect, once it is correlated with other important socioeconomic indicators that are not present in the model, as the mean income. Nevertheless, the presence of sanitation is largely recognized as a basic preventive factor against a range of diseases that afflict, mainly, the child population (see Neri 2007). The *Gini* index and the indicator of piped water in turn, don´t present statistical significance. It may be partially because these variables don´t hold enough variance between the municipalities. Moreover, the correlation amongst the variables at second level might undermine their real “explanatory power”.

[29]Because the mean income is highly correlated with schooling, we had econometric difficulties to estimate both together in the same model. In this sense, we decided to explore only the second indicator.

Lastly, the health family´s program reveals a favorable impact on subjective health. The governmental health assistance reports a significant positive signal, suggesting its effective role on the population health´s improvement, especially for the worst-off.

Although the statistical significance of some contextual covariates, it seems that our municipal-level measures are not well fitted to explain, even partially, the second level variance, which even increased 0.016 after the inclusion of the group-level covariates[30]. Possibly, the quantitative indicators we used are misspecified to capture more subtle and complex contextual socioeconomic aspects of a whole municipality, which is not an easy task, given the limitation of our dataset for this sort of analysis.

It can be argued that the variance between the municipalities may be somehow a result of higher geography entities. In fact, institutional, political and cultural/historical frameworks of the states are dimensions commonly shared by their population, and can be related to the health infra structure available along with cultural habits and lifestyles. In this sense, we extent our final model to a third level, which corresponds to 18 Federal States of Brazil (given our reduced sample). As the contextual variables available didn't explain any second level variation, we compute the last model only controlling for the individuals variables.

The three-level random effects in the last column of Table 2.4 provide an interesting result. The correlation coefficient between the municipalities decreased from 9% to 4%. On the other hand, 11% of the total variance lies between the states. This means that a portion of the variance between municipalities is further explained by differences in the states where they are clustered. Indeed, the population of the Brazilian federal states present significant differences not only in terms of socioeconomic circumstances, cultural behaviors and diet habits; but also, provide distinct security and health policies in both qualitative and availability aspects. Not to mention the climate and latitude/longitude differences. Still, it is pertinent to consider the possibility that traditions, values, cultural and lifestyle differences across the main regions in Brazil, may lead people to evaluate their health through different cut-off points, possibly generating as a result, "cultural heterogeneity bias".

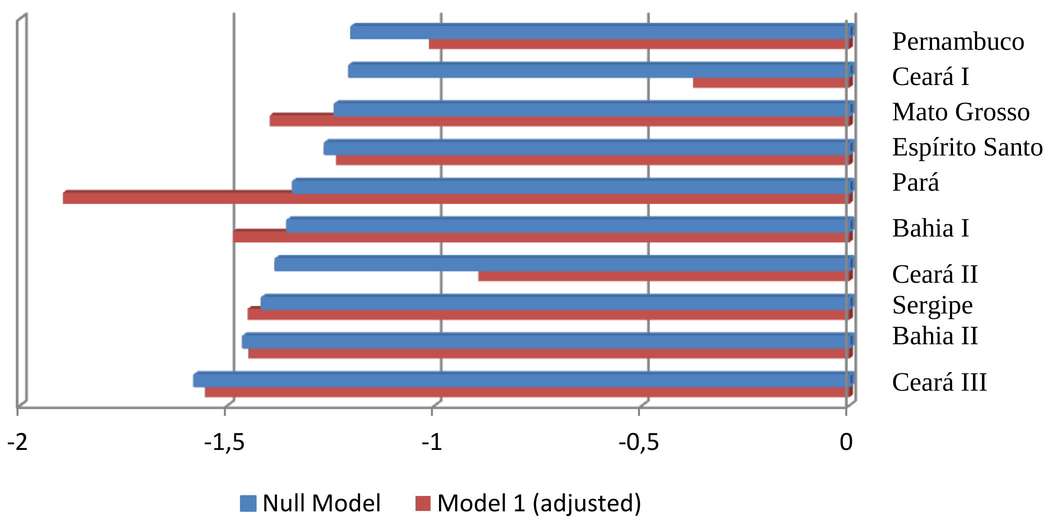
[30]As pointed out by Stafford (2001), the addition of explanatory variables at the second level of hierarchical analysis, may produce instead, an increasing in the variance, indicating that there are greater differences between leve-2 entities than 'would be expected on the basis of the variables added'.

2.5.1 Residuals Analysis

We lastly explore the second level residuals replicating a sort of analysis employed by many studies (e.g. Stafford et. al, 2001; Wiggins et. al., 2002). The idea relies on ranking the area units according to the regression´s residuals in order to provide where the municipalities with the most positive and negative residuals are located. Put differently, these evidences suggest that the municipalities which the population reported better health than the overall mean predicted by the multilevel model offer contextual features more favorable to subjective health. On the other hand, those municipalities with more poor health than the predicted, are considered "less conducive" to better health. We will refer to them respectively as the "healthy" and "unhealthy" municipalities.

Since we are not able to recognize the municipalities, we identify them by the Federal States which they are situated, and correspond to the figures´ vertical axis. We can identify as well, what happens with the outlier residuals as we adjust the model with individual characteristics. Figure 2.3 and 2.4 plots the first ten municipalities that obtained respectively the most negative and positive residuals referent to the null model.

Figure 2.3 - The ten "unhealthiest" municipalities predicted by the null and individual adjusted Multilevel Models



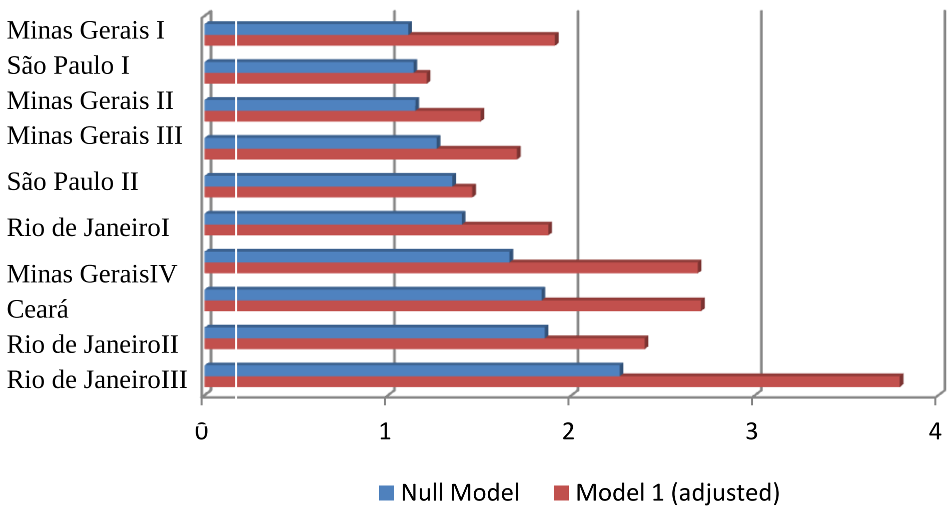
Source: PNAD 2008. Self Elaborated
Note 1: The bars correspond to the municipalities, identified by their states

The bars show the magnitude of the residuals. As we can note through Figure 2.3, the null model presents for the majority municipalities, larger residuals than the adjusted model

(Model 1), as one would expected *a priory*. From a general perspective, we observe the existence of "response" heterogeneity between the municipalities when we compare the "behavior" of the residuals as we move from the null to the adjusted model. For example, despite of owning very similar residuals in the null model, the municipalities from Pernambuco, "Ceará I" and Mato Grosso, present great distance in their residuals after individual control.

It is also possible to verify if there exists any cluster of "healthy" or "unhealthy" municipalities within a unique state; or conversely, if they don´t follow any regional pattern, and in fact, are spread out across the country. Regarding the unhealthy municipalities, those that present the largest magnitudes are respectively from Ceará III, Bahia II and Sergipe. Indeed, the most municipalities are from Northeast, whereas there is only one from southeast and none from south´s states.

Figure 2.4 - The ten "healthiest" municipalities predicted by the null and individual adjusted Multilevel Models



Source: PNAD 2008. Self Elaborated

On the other hand, as reported by Figure 2.4, among the "healthiest" municipalities, 90% are from southeast: three from Rio de Janeiro; four from Minas Gerais; and two from the State of São Paulo. Other evidence that draws attention is the presence of a municipality of Ceará. The State which had three municipalities amongst the most "unhealthy" ones, also detain a city with large positive SAH residuals. This may indicate the presence of considerable contextual health variations within a same state; supporting the observed second-level variance.

Furthermore, the residuals pattern from the adjusted model provides a contra-intuitive result; i.e., after individual control, the residuals became larger for this group of municipalities. A more accurate explanation for such evidence would need more detailed investigation about the population and the physical and social features of these localities.

2.6 CONCLUSIONS

The main purpose of this paper relies on estimating in what extent the variability of Self Assessed Health in Brazil is a result of the municipality where the people are inserted in. The recommended multilevel model technique, adapted for dependent ordinal variables, was employed in order to appropriately separate out the compositional, from the contextual effects, over the perceived health.

We took advantage of the available sample of municipalities in PNAD (2008) database, and employed them as our second level unities. We additionally extended the last hierarchical model to a third level, corresponding to the federal states, in order to further analyze if the contextual effects found in our previews models, may be also a result of non observed differences between the States.

Our main findings showed a correlation coefficient of 7% after composition control, indicating a systematic variance in SAH between the municipalities that cannot be explained by the differing individual composition of these localities. This evidence suggests that municipal variations in perceived health were not fully explained by individual socioeconomic circumstances, remaining statistic significant variability; possibly in virtue of not observed contextual features. In this sense, we tried by means of a set of aggregated variables at municipal level, to explain at least part of this unobserved variance. However, the results showed that the macro-level measures we adopted were unable to further explain any variance between the municipalities.

Furthermore, when we computed the federal states as a third scale of SAH variance, an interesting result arose: 11% of the total SAH variance lied between the states, and it still remained 4% between the municipalities. This evidence might suggest the great diversity across the states of the country in terms of socioeconomic conditions, as variations in quality and availability of the health services; sanitation coverage; political and institutional frameworks; cultural behaviors etc.; that in turn, may reflect in how the population perceives their own health. Still, the procedure of residuals´ analysis reported that most of the municipalities which presented the highest positive unexplained proportion of SAH after prediction, were located in the southeast´s states. On the contrary, those municipalities with the most negative residuals lied in their majority, in northeast´s states.

Given the generality of our analysis and the complexity of the contextual effects, we were not able to point out any specific municipal feature or chain that might affect directly their population´s health. Nevertheless, the evidences found here suggested that there exist contextual effects in Brazil over and above individual circumstances. Furthermore, the evidences provided empirical results never explored in the country, and encourage investment on surveys in Brazil that capture health information from populations belonging smaller geographic units, in order to identify important local social and physic features that might promote or damage their inhabitant´s health.

We believe that further studies in Brazil that take into account particular demographic profiles in the contextual effect analysis are also necessary for policy focalization. For example, it is reasonable to expect that the children´s health status hold a different "response-function" against environmental risks if compared with the adults, once they are more exposed to surrounding hazards, as lack of sewage.

Finally, the results we found here, highlights the potential of public policy to promote the health of their inhabitants in the sense of not focusing only on individual socioeconomic circumstances, but rather, improving the physical and social environments where they live.

Examples rely on providing an unbiased service of sewage and garbage removal and creating opportunities that favor and encourage – especially the worst-off – the adoption of healthier behaviors and lifestyles, such as offering spaces for recreation and urban facilities.

REFERENCES

ADLER, N.E.; OSTROVE, J.M. Socioeconomic Status and Health: What We Know and What We Don't, in ADLER, N.E.; MARMOT, M. G.; MCEWEN, B. S.; STEWART, J. *Annals of New York Academy of Sciences*: Socioeconomic Status and Health in Industrial Nations: Social, Psychological, and Biological Pathways, vol. 896, 3-15, 1999.

ALLISON, R.A.; FOSTER. J.E. Measuring Health Inequality Using Qualitative Data. *Journal of Health Economics*, 23, 505–524, 2004.

ANAND, S.; PETER, F.; SEN, A. Public Health, Ethics, and Equity, Oxford University Press, New York, 2004.

APOUEY, B. Measuring health polarization with self-assessed health data. *Health Economics*, vol. 16, issue 9, 2007.

BARROS, R.; CARVALHO, M.; FRANCO, S.; MENDONÇA R. A Queda Recente da Desigualdade de Renda no Brasil, Brasília, vol. 2, IPEA, 2007.

BARUFI, A.M; HADDAD, E.; PÁEZ, A. Infant Mortality in Brazil.1980-2000: A spatial panel data analysis, *BMC Public Health*, 12:181, 2012.

BELLÙ, L. Inequality and Axioms for its Measurement. *Agricultural Policy Support. Policy Assistance Division*, FAO, Italy, 2006.

BLAKELY, T. A; WOODWARD, A J. Ecological effects in multi-level studies. *Journal of epidemiology and community health*, *54*(5), 367-74, 2000.

BRASIL. MINISTÉRIO DA SAÚDE. Política Nacional de Atenção Integral à Saúde da Mulher: Princípios e Diretrizes. *Secretaria de Atenção à Saúde Departamento de Ações Programáticas Estratégicas*, 2004.

____________________________ As causas sociais das iniquidades em saúde no Brasil. *Relatório Final da Comissão Nacional sobre Determinantes Sociais da Saúde*. Rio de Janeiro, Ed. Fiocruz, 2008.

BRYK A, RADENBUSH S. Hierarchical linear models: applications and data analysis methods. Newbury Park: Sage Goldstein H. Multilevel statistical models. London, 1992.

CROUCHLEY, R.; STOTT, D.; PRITCHARD, J. Multivariate Generalized Linear Mixed Models via sabre Stata, Version 1, 2009.

CURTIS, S.; REES JONES, I. Is there a place for geography in the analysis of health inequality? *Sociology of Health & Illness*, 20(5), 645–672, 1998.

DANIELS, N.; KENNEDY, B.; KAWACHI, I. In ANAND, S.; PETER, F.; SEN, A. *Public Health, Ethics, and Equity*, Oxford University Press, New York, 2004.

DIEZ-ROUX A. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American Journal of Public Health*; 88:216–22, 1998.

DIEZ-ROUX A. Next steps in understanding the multilevel determinants of health. *Journal of Epidemiology Community Health*; 62:957–9, 2008.

DIEZ-ROUX, A.; NIETO, F.; MUNTANER, C. et al. Neighborhood environments and coronary heart disease (a multilevel analysis). *American Journal of Epidemiology*, 146:48–63, 1997.

DUNCAN, C., JONES, K., MOON, G. Do places matter? A multi-level analysis of regional variations in health-related behavior in Britain. *Social science & medicine*, *37*(6), 725-33, 1993.

DUNCAN, C., JONES, K., MOON, G. Context, composition and heterogeneity: using multilevel models in health research. *Social science and medicine* 46(1), 97-117, 1998.

DUNCAN, C., JONES, K., MOON, G. Psychiatric morbidity: a multilevel approach to regional variations in the UK. *Journal of epidemiology and community health*, *49*(3), 290-5, 1995.

ECOB, R.; MACINTYRE, S. Small area variations in health related behaviors: do these depend on the behavior itself, its measurement, or on personal characteristics? *Journal of Health and Place*, 6, 261–274, 2000.

ELLAWAY, A. and MACINTYRE, S. Neighborhoods and health. In: BROWN, T., MCLAFFERTY, S. and MOON, G. (eds.). A Companion to Health and Medical Geography. Blackwell Publishing, Oxford, UK, 2009.

ELSTAD, J. I. Does the socioeconomic context explain both mortality an income inequality?Prospective register-based study of Norwegian regions. *International journal for equity in health*, *10*(1), 7, 2011.

FOGUEL, M. N.; BARROS . R. P. The Effects of Conditional Cash Transfer Programs on Adult Labor Supply: An Empirical Analysis Using a Time-Series-Cross-Section Sample of Brazilian Municipalities. *Estudos Econômicos*, Sao Paulo, v. 40, n. 2, 2010.

GERDTHAM, U. Income-related inequality in life-years and quality-adjusted life-years. *Journal of Health Economics* 19:6, 1007-1026, 2000.

GOLDSTEIN, H. Multilevel Mixed linear Model Analysis using iterative generalized least squares. *Biometrika*; 73:43-56, 1986.

GÓMEZ, P.; NICOLÁS, A. Regional differences in socioeconomic health inequalities in Spain. Departament d'Economia i Empresa and CRES, Universitat Pompeu Fabra, Barcelona, Spain, 2004.

GROOT, W. Adaptation and scale of reference bias in self-assessment of quality of life. *Journal of Health Economics*; 403-20, 2000.

GWATKIN, D.; RUTSTEIN, S.; JOHNSON, K.; PANDE, R.; WAGSTAFF, A. Socioeconomic Differences in Health, Nutrition and Population. Washington DC, The World Bank, 2000.

HAAN M, KAPLAN G.A.; CAMACHO T. Poverty and health: prospective evidence from the Alameda County Study. *American Journal of Epidemiology*, 125:989-98, 1987.

HABIBOV, N. N.; AFANDI, E. N. Self-rated health and social capital in transitional countries: Multilevel analysis of comparative surveys in Armenia, Azerbaijan, and Georgia. *Social science and medicine 72*(7), 1193-204, 2011.

HAN, S., KIM, H., LEE, H. A multilevel analysis of social capital and self-reported health: evidence from Seoul, South Korea. *International journal for equity in health*, *11*(1), 3, 2012.

HART, C.; ECOB, R.; SMITH, G. D. People, places and coronary heart disease risk factors: a multilevel analysis of the Scottish Heart Health Study archive. *Social science & medicine*, 45(6), 893-902, 1997.

HUMPHREYS, K.,; CARR-HILL, R. Area variations in health outcomes: artifact or ecology. *International journal of epidemiology*, *20*(1), 251-8, 1991.

HUMPHRIES, K.H; VAN DOORSLAER, E. Income-related health inequality in Canada. *Social Science and Medicine*, vol.50(5), 2000.

IDLER, E.L.; BENYAMINI, Y. Self-rated health and mortality: a review of twenty seven community studies. *Journal of Health and Social Behavior*, 38:21-37, 1997.

JONES, K., & DUNCAN, C. Individuals and their ecologies: analyzing the geography of chronic illness within a multilevel modeling framework. *Health and Place*, *1*(1), 27-40, 1995.

JÜRGES, H. True health *vs* response styles: Exploring cross-country differences in self-reported health. *Health Economics*, 16, 163–178, 2007.

KAKWANI, N.; WAGSTAFF, A.; VAN DOORSLAER, E. Socioeconomic inequalities in health: measurement, computation, and statistical inference. *Journal of Econometrics*, 77 (1), 1997.

KAPLAN G. People and places: contrasting perspectives on the association between social class and health. *International Journal of Health Service*. ;26:507-19, 1996.

KAWACHI, I.; SUBRAMANIAN, S; ALMEIDA-FILHO, N.A. Glossary for health inequalities. *Journal of Epidemiology and Community Health*; 56:647–652, 2002.

KENNEDY, B. P., KAWACHI, I., GLASS, R.; PROTHROW-STITH, D. Income distribution, socioeconomic status, and self rated health in the United States: multilevel analysis. *BMJ* 317(7163), 917-21. 1998.

KERKHOFS, M. and LINDEBOOM, M. Subjective health measures and state dependent reporting errors. *Health Economics*, 4: 221–235, 1995.

KIM, D.; KAWACHI, I. A multilevel analysis of key forms of community and individual-level social capital as predictors of self-rated health in the United States. *Journal of urban health* : bulletin of the New York Academy of Medicine, *83*(5), 813-26, 2006.

LAURELL, A.C. La salud-enfermedad como proceso social. *Revista Latinoamericana de Salud*, N° 2, Mexico, 1982.

LAURIDSEN, J.; CHRISTIANSEN, T.; HAKKINEN, U. Measuring inequality in self-reported health discussion of a recently suggested approach using Finnish data, *Health Economics*, vol. 13, 725–732, 2004.

LINDEBOOM, M.; DOORSLAER, E. Cut point Shift and Index Shift in Self Reported Health. *Discussion Paper N°.1286*, Bonn, Germany, 2004.

MACINTYRE, S., ELLAWAY, A., CUMMINS, S. Place effects on health: how can we conceptualize, operationalize and measure them? *Social Science and Medicine*, *55*(1), 125-139, 2002.

MACINTYRE, S.; MACIVER, S.; SOOMANS, A. Area, class and health: should we be focusing on places or people, *Journal of Social Policy*, 22, 2, 213–34, 1993.

MACINTYRE, S.; SOOMANS, A. Health and perceptions of the local environment in socially contrasting neighborhoods in Glasgow, *Health and Place*, 1, 1,15–26, 1995.

MADDEN, D. Ordinal and cardinal measures of health inequality: an empirical comparison, *Health Economics*, 19: 243–250. 2008.

MCCALL, D.; NAMBA, N.; FRIEDMAN M. A Healthy Future. U.S. Department of State, Bureau of International Information Programs. 2011.

MITCHELL, R., GLEAVE, S., BARTLEY, M., WIGGINS, D., JOSHI, H. Do attitude and area influence health? A multilevel approach to health inequalities. *Health and Place*, *6*(2), 67-79. 2000.

NERI M., Trata Brasil – Saneamento e Saúde. FGV/IBRE, CPS, 2007.

NOGUEIRA, H. A saúde e os lugares. Coimbra, Imprensa da Universidade de Coimbra, Lisboa, 1ª Edição, 2008.

NORONHA, K.V.M.S., ANDRADE, M.V. O Efeito da Distribuição de Renda Sobre o Estado de Saúde Individual no Brasil. Anais da ANPEC, 2005.

NORONHA, K.V.M.S., ANDRADE, M.V. Desigualdades sociais em saúde: evidências empíricas sobre o caso Brasileiro. Texto para Discussão n°.171, 2002.

O'DONNELL, O.; VAN DOORSLAER, E.; WAGSTAFF, A.; LINDELOW, M. Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation. Washington, DC: World Bank Publications, 2007.

O'CONNELL, A.A. An illustration of multilevel models for ordinal response data. In: Proceedings of the 8th International Conference on Teaching Statistics, International Statistics Institute, Ljubljana, Slovenia, 2010.

PHILLIMORE, P.,; READING, R. A rural advantage? Urban-rural health differences in northern England. *Journal of public health medicine*, *14*(3), 290-9, 1992.

PICKETT, K. E.; PEARL, M. Multilevel analyses of neighborhood socioeconomic context and health outcomes: a critical review. *Journal of epidemiology and community health*, *55*(2), 111-22, 2001.

RABE-HESKETH, S. Multilevel modeling of ordered and unordered categorical responses. Institute of child Health, London, 1-12, 2003.

RABE-HESKETH, S., SKRONDAL, A. Multilevel and Longitudinal Modeling Using Stata. Third Edition, Vol. I: Continuous Responses. College Station, TX: Stata Press, 2012.

RABE-HESKETH, S., SKRONDAL, A., PICKLES, A. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, *128*(2), 301-323, 2005.

RABE-HESKETH, S., SKRONDAL, A., PICKLES, A. Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata Journal* 2: 1–21, 2002.

RICE N, JONES A. Multilevel models and health economics. *Health Economics*, 6, 561-75, 1997.

RISTE L, KHAN F, CRUICKSHANK K. High prevalence of type 2 diabetes in all ethnic groups, including Europeans, in a British inner city: relative poverty, history, inactivity, or 21st century Europe? *Diabetes Care*. University of Manchester Medical School, UK. 24(8):1377-83, 2001.

ROWLEY D.L; HOGAN V. Disparities in Infant Mortality and Effective, Equitable Care: Are Infants Suffering from Benign Neglect? *Annual Review of Public Health*, (33)75 -87, 2012.

SACOSS. South Australian Council of Social Service. Social Determinants of Health. *Sacoss Information Paper*, Australia, 2008.

SÁNCHEZ-VAZNAUGH, E.V.; KAWACHI, I. SUBRAMANIAN SV, SÁNCHEZ BN, ACEVEDO-GARCIA D. Do socioeconomic gradients in body mass index vary by race/ethnicity, gender, and birthplace? *American Journal of Epidemiology*, 169(9), 2009.

SEN, G.; ÖSTLIN, P.; GEORGE, A. Unequal, Unfair, Ineffective and Inefficient. Gender Inequity in Health: Why It Exists and How We Can Change It. Final Report to the WHO Commission on Social Determinants of Health. Women and Gender Equity Knowledge Network, 2007.

SHMUIELI, A. Socio economic and demographic variation in health and in its measures: the issue of reporting heterogeneity, *Social Science and Medicine*, vol. 57, 125-134, 2003.

SILVA, P.; PESSOA, M.; LILA, F. Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral. *Ciência e Saúde Coletiva*, 7(4): 659-670, 2002.

SLOGGETT, A.; JOSHI, H. Higher mortality in deprived areas: community or personal disadvantage? *BMJ*, 1470-4, 1994.

SOARES, R. Health and the evolution of welfare across Brazilian municipalities. NBER *Working* Paper 13087, Cambridge, 2007.

SOUSA, M. H. SILVA, N. Estimativas obtidas de um levantamento complexo. *Revista de Saúde Pública*, vol.37, no.5, p.662-670, 2003.

SOUTH AUSTRALIAN COUNCIL OF SOCIAL SERVICE (SACOSS). Social Determinants of Health: Sacoss Information Paper, Australia, 2008.

STAFFORD, M.; BARTLEY, M.; MITCHELL, R.; MARMOT, M. Characteristics of individuals and characteristics of areas: investigating their influence on health in the Whitehall II study. *Health and place*, 7(2), 117-29, 2001.

THE WORLD BANK. Health, Nutrition, and Population Sector Strategy Paper, Washington (D.C), 1997.

TUBEUF, S.; PERRONNIN, M. New prospects in the analysis of inequalities in health: a measurement of health encompassing several dimensions of health. HEDG W/P 08/01 University of York, 2008.

VAN DOORSLAER E., GERDTHAM, U. Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. Social Science and Medicine,57/9: 1621–9, 2003.

VAN DOORSLAER, E.; KOOLMAN, X. Explaining the differences in income-related health inequalities across European countries. *Health Economics*, 13(7), 2004.

VAN DOORSLAER, E.; WAGSTAFF, A.; BLEICHRODT, H. Income-related inequalities in health: Some international comparisons. *Journal of Health Economics*, vol. 16, 93-112, 1997.

VAN DOORSLAER, E.Y; JONES, A. Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics*, Issue 1, 61-78, 2002.

WAGSTAFF, A. Socioeconomic inequalities in child mortality: comparisons across nine developing countries. *Bulletin of the World Health Organization*, N° 78, 2000.

WAGSTAFF, A.; PACI, P.; JOSHI, H. Causes of Inequality in Health: Who are you? Where you live? Or who your parents? World Bank. *Policy Research Working Paper*. N°. 2713. Washington (D.C), 2001.

WAGSTAFF, A.; VAN DOORSLAER, E. Measuring inequalities in health in the presence of multiple category morbidity indicators. *Health Economics*, vol. 3, 281–291, 1994.

WIGGINS, R. D., JOSHI, H., BARTLEY, M., GLEAVE, S., LYNCH, K., & CULLIS, A. Place and personal circumstances in a multilevel account, *Social Science and Medicine 54*(5), 827-838, 2002.

WIGHT, R. G., CUMMINGS, J. R., MILLER-MARTINEZ, D., KARLAMANGLA, A. S., SEEMAN, T. E., & ANESHENSEL, C. S. A multilevel analysis of urban neighborhood socioeconomic disadvantage and health in late life. *Social science and medicine*, *66*(4), 862-72, 2008.

WORLD BANK. Health, Nutrition, and Population Sector Strategy Paper, Washington (D.C), 1997.

WORLD HEALTH ORGANISATION. 'Declaration of Alma-Ata' International Conference on Primary Health Care, Alma-Ata, USSR, 6-12, 1978.

WORLD HEALTH ORGANIZATION. Environmental Health Inequalities in Europe, The WHO European Centre for Environment and Health, Denmark, 2012.

YEN, I. H.;KAPLAN, G.A. Poverty area residence and changes in depression and perceived health status: evidence from the Alameda County Study. *International Journal of Epidemiology*, 28, 1999.

YIENGPRUGSAWAN, V.; LIM, L.L.; CARMICHAEL, G.A.; DEAR, K.; SLEIGH, A. Decomposing socioeconomic inequality for binary health outcomes: an improved estimations that does not vary by choice of reference group. *BMC Reaserch Notes*, Australia, 3:57 2010.

ANNEX

Table 1.11 - Independent Variables of the Probit Model: Descriptive Statistics for the year 2003

VARIABLES	*code*	Description	Mean	SD	nº of obs.
Gender	*gender*	Men=1	0.486	0.500	237881
Age	*age*	The age variable ranges from 16 to 65 years old	35.404	13.357	237881
Race	*race*	Dummy representing white=1 and non-white=0	0.493	0.500	237881
lnincome	*ln_pcincome*	Per capita household´ income in log scale. Min.= 0.847 and Max=11.918	5.392	1.078	237881
Years of schooling	*schooling*	Number of years of schooling. Ranges from 1 year or less to 16 years or more	8.100	4.305	237881
Married	*married*	Couple has one child or more =1 (Dummy)	0.760	0.427	237878
Couple with children	*married_child*	Couple has no child = 1 (Dummy)	0.247	0.431	237881
Couple without children	*married_nochild*	Mother with both children younger and/or older than 14 years old (Dummy)	0.110	0.313	237881
Single mother	*single mother*	The individual is not married neither has child = 1 (Dummy)	0.154	0.361	237881
Single mother 1	*mother1*	The individual is economically active and worked in the week of reference = 1 (Dummy)	0.029	0.167	237878
Single mother 2	*mother 2*	The individual is economically active but didn´t work in the week of reference = 1 (Dummy)	0.101	0.301	237881
Single (no child)	*single*	The individual holds private health insurance = 1 (Dummy)	0.087	0.281	237881
Worked in the week of reference	*employed*	Urban area=1; rural=0	0.645	0.478	237866
unemployed	*unemployed*	Metropolitan region=1 ; non metropolitan region = 0	0.092	0.289	237866
Health insurance	*health_ins*	Men=1	0.256	0.437	237867
urban_rural areas	*urb*	The age variable ranges from 16 to 65 years old	0.864	0.343	237881
Metropolitan Region	*reg_metrop*	Dummy representing white=1 and non-white=0	0.390	0.488	237881

Source:PNAD 2008

Table 1.12 - Independent Variables of the Probit Model: Descriptive Statistics for the year 1998.

VARIABLES	*code*	Description	Mean	SD	nº of obs.
Gender	*gender*	Men=1	0.4854	0.4998	206070
Age	*age*	The age variable ranges from 16 to 65 years old	34.9918	13.3208	206070
Race	*race*	Dummy representing white=1 and non-white=0	0.5296	0.4991	206070
lnincome	*ln_pcincome*	Per capita household´ income in log scale. Min.= 0.847 and Max=11.918	5.0408	1.0968	206070
Years of schooling	*schooling*	Number of years of schooling. Ranges from 1 year or less to 16 years or more	7.3613	4.2686	206070
Married	*married*	Couple has one child or more =1 (Dummy)	0.7803	0.4140	206058
Couple with children	*married_child*	Couple has no child = 1 (Dummy)	0.6814	0.4659	206058
Couple without children	*married_nochild*	Mother with both children younger and/or older than 14 years old (Dummy)	0.0989	0.2985	206070
Single mother	*single mother*	The individual is not married neither has child = 1 (Dummy)	0.1413	0.3483	206070
Single mother 1	*mother1*	The individual is economically active and worked in the week of reference = 1 (Dummy)	0.0263	0.1599	206058
Single mother 2	*mother 2*	The individual is economically active but didn´t work in the week of reference = 1 (Dummy)	0.0887	0.2843	206070
Single (no child)	*single*	The individual holds private health insurance = 1 (Dummy)	0.0784	0.2688	206070
Worked in the week of reference	*employed*	Urban area=1; rural=0	0.6399	0.4800	206055
unemployed	*unemployed*	Metropolitan region=1 ; non metropolitan region = 0	0.0830	0.2759	206035
Health insurance	*health_ins*	Men=1	0.2671	0.4424	206054
urban_rural areas	*urb*	The age variable ranges from 16 to 65 years old	0.8338	0.3723	206070
Metropolitan Region	*reg_metrop*	Dummy representing white=1 and non-white=0	0.4146	0.4927	206070

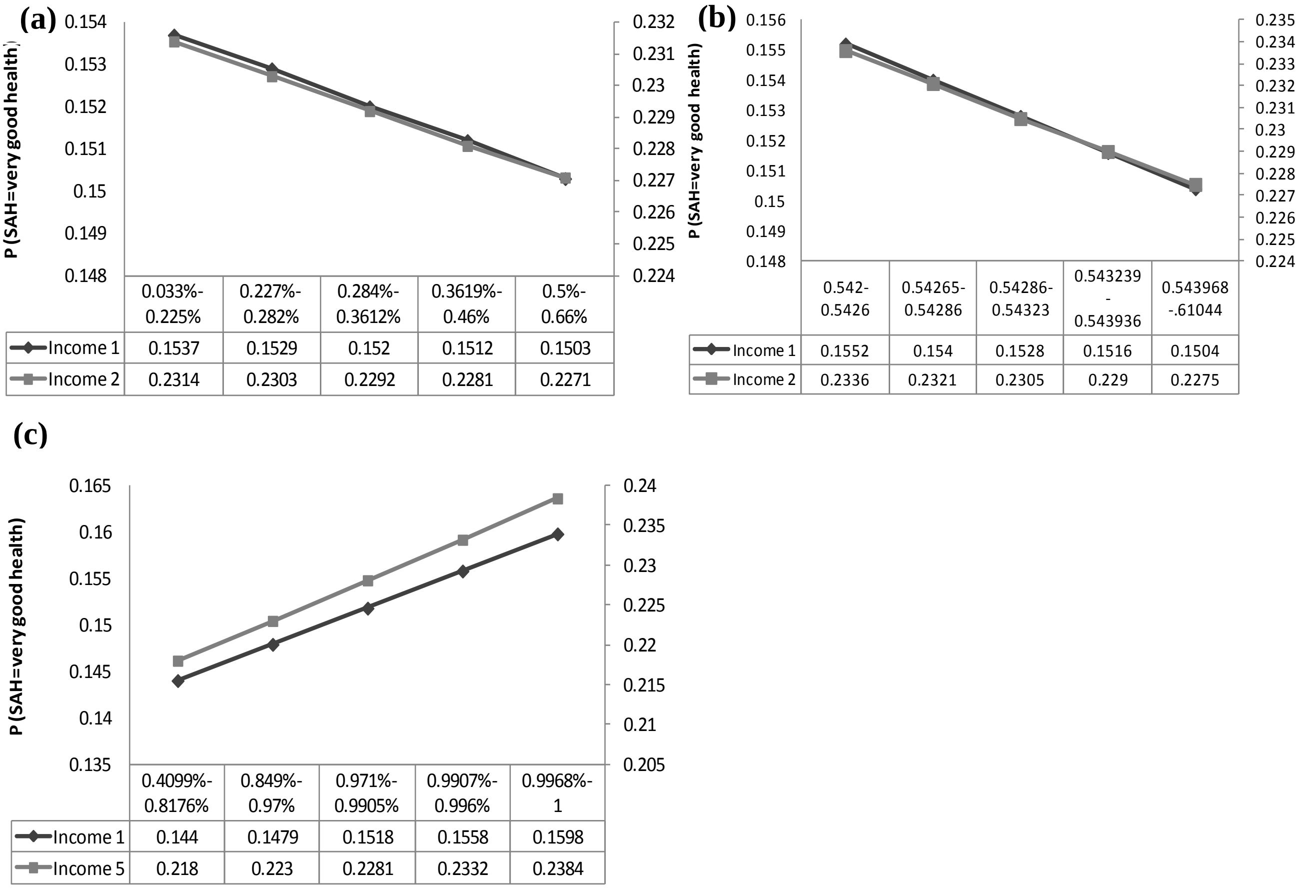
Source: PNAD 2008

Table 2.5 - Correlation Matrix of Contextual Variables

	mn_income	*lowschooling*	*Gini*	*%Sewage*	*%Water*	*%family_prog*
mn_income	1					
lowschooling	-0.5719	1				
Gini	-0.0644	0.2793	1			
%Sewage	0.485	-0.3975	-0.0886	1		
%Water	0.434	-0.3385	-0.073	0.2505	1	
%family_program	-0.3237	0.3625	0.0677	-0.335	-0.2402	1

Source: PNAD 2008

Figure 2.5 - Probability of answering "very good health" of the worst off – income 1 – and better-off – income 5 – according to the quintiles groups of municipalities ranked into: (a) Proportion of adults over 25 years old with less than six years of schooling; (b) Gini Coefficient; (c) proportion of households with piped water



Source: PNAD 2008. Self Elaborated