

# Mahalanobis' distance and propensity score to construct a controlled matched group in a Brazilian study of health promotion and social determinants

*Distância de Mahalanobis e escore de propensão para seleção de amostra pareada de controle em um estudo Brasileiro de promoção da saúde e determinantes sociais*

Valéria Troncoso Baltar<sup>I</sup>, Clóvis Arlindo de Sousa<sup>II</sup>, Marcia Faria Westphal<sup>III</sup>

**ABSTRACT:** In observational epidemiology it is usual to select a control group to study the effects of certain exposures on human health. Intervention studies are well known among epidemiologists but it is not very frequent in other areas of research. In this paper we propose the same idea of intervention studies and the use of three methods for a health promotion research control group selection: Propensity score, Mahalanobis' distance and Mahalanobis within Propensity Calipers. In the original project, "Health and Local Development: a progress review towards the millennium goals with relation to health in the Brazilian cities which develop social agendas", cities with social agendas from Brazil were matched separately by state. In the state of Paraná there are 397 cities. Of these, 34 presented social agendas implemented and active since, at least, 2004. Five variables measured in 2000 were considered for the matching: population size, human development index of income, human development index of education, percentage of literacy and vaccine coverage. As a result, among these three methods, the Mahalanobis by itself was considered the less efficient. In conclusion, the propensity, which is a very simple linear score, presented very good matched sample. However, the Mahalanobis within Calipers was the method that provided the best result.

**Keywords:** Matched-pair analysis. Control groups. Mahalanobis' distance. Propensity score. Caliper matching. Epidemiology.

<sup>I</sup>Epidemiology and Biostatistics Department, Institute of Community Health, *Universidade Federal Fluminense* – Rio de Janeiro (RJ), Brasil.

<sup>II</sup>Physical Education and Medicine Department, *Universidade Regional de Blumenau* – Blumenau (SC), Brasil.

<sup>III</sup>Practice in Public Health Department, School of Public Health, *Universidade de São Paulo* – São Paulo (SP), Brasil.

**Corresponding author:** Valéria Troncoso Baltar. Departamento de Epidemiologia e Bioestatística, Universidade Federal Fluminense. Travessa Marquês de Paraná, 303 / 3º andar, Centro, CEP 24030210, Niterói, RJ, Brasil. Email: vtbaltar@id.uff.br

**Conflict of interests:** nothing to declare – **Financing Source:** National Council for Scientific and Technological Development (CNPq - process number 409821/2006-3).

**RESUMO:** Em epidemiologia observacional, é frequente o uso de grupos controle para avaliação do efeito de variáveis de exposição em desfechos na saúde de pessoas, porém este método não é muito utilizado em outras áreas. Este artigo propõe a aplicação da ideia de estudos de intervenção, com base em seleção de grupo controle, utilizando três métodos de seleção de amostra (escore de propensão, distância de Mahalanobis e distância de Mahalanobis dentro da margem estabelecida pelo escore de propensão) para pesquisa de promoção da saúde. No projeto “Saúde e desenvolvimento local: análise dos progressos em relação aos objetivos de desenvolvimento do milênio relacionados à saúde, nas cidades brasileiras que desenvolvem agendas sociais”, cidades com agendas sociais foram pareadas com amostra controle sem agendas sociais, para cada um dos estados do Brasil. Neste artigo foi considerado o estado do Paraná que tem 397 cidades sendo 34 com agendas sociais implementadas desde pelo menos 2004. Cinco variáveis, coletadas em 2000, foram consideradas para o pareamento: tamanho populacional, índice de desenvolvimento humano econômico e educacional, percentual de pessoas escolarizadas e cobertura vacinal. O resultado do pareamento com o uso da distância de Mahalanobis foi o que apresentou menor qualidade. Conclui-se que o método do escore de propensão, o mais simples e mais facilmente utilizado, apresentou como resultado um grupo de controle confiável. Entretanto, a distância de Mahalanobis dentro de margens do escore de propensão é o método que obteve o melhor resultado.

**Palavras-chave:** Análise por pareamento. Grupos controle. Distância de Mahalanobis. Pontuação de propensão. Análise de pareamento por margem de propensão. Epidemiologia.

## INTRODUCTION

In observational research the central problem is that interventions (exposed) and control groups, that will be compared, may not be comparable prior to exposed intervention, so differences in outcomes may not represent a causal effect<sup>1</sup>. Observed pre-intervention differences need to be controlled by model adjustment or by matching. In fact, a random sampling of a non-exposed group would not be a good idea if the non-exposed sample is different from the exposed group regarding to background variables<sup>2</sup>.

Miettinen<sup>3</sup> proposed for the first time the use of a score based on several co-variables to adjust for confounding. Rosenbaum and Rubin<sup>4</sup> introduced the propensity score concept to control selection bias in cohort. Originally, the method was proposed to control imbalance among co-variables distributions for observational studies, in situations in which randomization was not possible.

Amongst a variety of controlled candidates (or non-exposed) we look for a sample which is comparable with the exposed sample to control selection bias. It is important to select specific co-variables that will be essential to guarantee that both groups are comparable. We suggest the application of a method recognized as efficient to estimate the effects of treatments and exposures on health outcomes in which the researcher looks for a representative sample of

controls to be matched with the treated or exposed. The fundamental principle is the maximum likelihood between groups considering the co-variables, excluding the response variable<sup>2</sup>.

At an individual level some strategies are used to select a control subject in order to reduce selection bias, and consequently cost. As an example, in case-control studies, controls can be selected in the neighborhood of cases, in areas close to the residence of the cases, providing a control with similar socio-economic and environmental conditions, with exception of the factor of interest. In intervention studies randomization can be used to provide similar groups to be treated or not. The same strategy can be applied to observational studies; in this case the municipalities' comparison can be made providing control cities to be matched to the treated or exposed ones<sup>5</sup>.

In the multicenter study "Health and Local Development: a progress review towards the Millennium Goals with relation to health in the Brazilian cities which develop social agendas" (*Saúde e desenvolvimento local: análise dos progressos em relação aos objetivos de desenvolvimento do milênio relacionados à saúde, nas cidades brasileiras que desenvolvem agendas sociais*)<sup>6</sup> some indexes are compared between the two groups (exposed and non-exposed to social agenda) throughout the years (since the exposed city had the agenda implemented). In this study Weighted Euclidian Distance was used to select control cities. The study was characterized as a retrospective cohort, the exposition and the outcome occurred before the beginning of the study. The matching control sample was essential not only in reducing bias but also in making the study feasible. The goal of the study was to evaluate the effect of social agendas on Millennium Development Goals (MDG). Several indexes were used as response variables, for example, the evaluation of the percentage of children less than a year old and with protein-energy malnutrition during some years after the agenda's implementation. A list of cities with social agendas all over Brazil was elaborated with basis on official websites and a committee participating in the project also made telephone contacts to make sure about the existence and duration of the agendas. To evaluate if these cities were having improvements due to the agendas, regarding a range of indexes considering MDG, the control group was necessary. Besides, the control group was necessary since many cities that were not included in the list with social agenda, could have some sort of social agenda implemented and should be excluded from the study. In the first phase of the study a range of control cities were selected, but among these cities it was necessary to ensure that they really do not had any agenda implemented. A second phase of the original study was established and each candidate city mayor's office was contacted to ensure these cities did not have social agendas implemented. The selection of a sample of possible candidates was essential since it is impossible to make sure all candidates do not have any agenda implemented.

The aim of this paper is to describe three different methods (Propensity score, Mahalanobis' distance and Mahalanobis within Propensity Calipers) to find and match this control group, providing bias control, with respect to background covariates. This matching made the main health promotion study feasible. Social agendas have been implemented in Brazil since 1990 with the main purpose of sustainable environmental development and promotion of local development. We considered as social agendas the "Agenda 21"<sup>7</sup>, "Cidades Saudáveis"<sup>8</sup> (Healthy

cities) and the national initiative Integrated and Sustainable Local Development (DLIS)<sup>9</sup>. In relation to the response variables on the main study, some characteristics were considered important for the matching of the cities. We considered population size (PS), human development index of income (HDII), human development index of education (HDIE), percentage of literacy (ALPHA) and vaccine coverage (VACC), all of which were collected for 2000.

We presented three methods of control group selection:

- Usual method of control group selection for comparison of treatments, the propensity score. This method was described in detail by Rosenbaum and Rubin<sup>1,4,10</sup> and D'Agostino<sup>11</sup>.
- The use of geometrical distance among the subjects (cities). A range of distances can be used, we selected Mahalanobis' distance<sup>12,13</sup>.
- The use of Calipers defined by propensity score and within Calipers the nearest Mahalanobis distance<sup>10</sup>.

## **METHODS**

### **STUDY SUBJECTS**

In the original project, "Health and Local Development: a progress review towards the millennium goals with relation to health in the Brazilian cities which develop social agendas"<sup>7</sup>, all the Brazilian cities were matched separately by state. To simplify, in the current article, we considered municipalities from Paraná state. In the state of Paraná there are 397 cities in which 34 presented social agendas implemented and active since, at least, 2004. This estate was selected since it does not contain a big capital or very complexes and too different cities that would need a more careful examination. We collected the range of municipalities' statistics for the year of 2000. The study was approved by the local ethic committee.

### **STATISTICAL METHODS**

We present a brief description of three matching methods. The propensity score, the Mahalanobis' matching distance and Mahalanobis matching within propensity Calipers as follows:

#### **Propensity score method**

Propensity score is a well known controls selecting method for non-randomized studies, with the aim of reducing bias<sup>4</sup>. The main purpose of this method is to select samples of controls to be compared with subjects under treatment in non-randomized studies. The control of bias has a co-variable basis selection. Studies which use the co-variables as

additional variables in the model usually do not have sufficient control for bias<sup>10</sup>; most of the time, because their sample sizes do not have enough statistical power.

Let's consider a binary random variable (intervention or exposure), assuming 0 or 1. Considering  $X_1, X_2, \dots, X_k$  as variables that are important for the matching, the propensity score is given by

$$\pi(x) = P(Y=1 / X_1=x_1, X_2=x_2, \dots, X_k=x_k). \quad (1)$$

This probability is estimated by a multiple logistic regression model. In this study, we adapted the same idea to select controls to match our exposed cities (with social agendas). In this regression the variables which are important for the matching will be included as co-variables. The researcher needs to include all the important variables for bias and confounding control with the exception of the outcome variable. In this study we considered as important factors those related to population size and development. We considered five co-variables: PS, HDII, HDIE, ALPHA and VACC (all of them measured in 2000).

The first step of this method is to perform a logistic regression with the entire data set, where the outcome is the exposed variable (with/without agenda), and the conditional probability of the agendas' occurrence is the propensity score. The score is used to match subjects from the sample with agendas with the sample without agendas. The smaller the difference between the scores, the more similar they are, so there is a pair matching. Therefore from the original sample, just a part of it will be kept for the final analysis. The subjects (cities) that will be dismissed are those which are more different from the exposed cities profile. The interpretation of the probability or score is the probability of the non-exposed (control) being exposed (in this case being a city with social agenda) allowing an almost-randomized study design. So, the score is the probability of the control being exposed (the control candidate being a city with a social agenda); similar to what is often carried out for randomized trials, where the pair guarantees that each subject has the same probability to receive a treatment or a placebo.

We performed the analysis for the entire Paraná state; and for each city with an agenda we selected two control candidates. Two controls per each city with social agenda are enough to explain the method in detail and it is a usual number for matching in epidemiology. In the original project we carried out the same analysis for each state in the country providing a list with 10 control candidates. Then a second phase of the project was implemented. Each control candidate was re-evaluated by the regional committee by checking if the controls were similar to their exposed cities evaluating other variables that were not possible to consider for the matching (qualitative analysis). The regional committee also checked if the control candidate was a real control, confirming if the city did not have social agendas, which would be impossible to know if the matching was not provided. In the end, many controls candidate could not be used, so a list of 10 was necessary in this particular situation, but it is not a typical situation.

We considered the following model:

$$\log \left\{ \frac{\pi(x)}{1-\pi(x)} \right\} = \alpha + \beta_1 PS + \beta_2 HDII + \beta_3 HDIE + \beta_4 ALPHA + \beta_5 VACC, \quad (2)$$

where the parameter  $\alpha$  and the vector of parameters  $\beta$  are estimated based on the whole sample.  $\pi(x)$  is the probability of a city having a social agenda ( $Y=1$ ), as given in expression (1).

The procedure is very simple: Cities with a social agenda are ranked randomly and the first is selected. The control candidate with the most similar score (in module) is selected. The control city is taken out from the range of candidates' sample. The second city with a social agenda is selected and also the control candidate with the most similar score. The second control city is excluded from the candidate's sample. If the study selects  $k$  ( $k > 1$ ) controls for each exposed, the algorithm needs to start again with the remaining range of candidate's sample (without reposition).

We carried out the procedure twice (because of simplicity, as mentioned before), considering that some of the controls would be excluded in phase 2, this number can be higher accordingly with the objectives of each study.

## Mahalanobis distance method

Geometric distances are well known in a varied range of sciences. They are used to analyze similarities between a pair of subjects. Cluster analysis can be a useful tool in summarizing results in any area of research, and it uses a specific distance selected by the researcher. For example, in principal components analysis it is possible to use the Weighted Euclidian distance between the subjects and then the most similar pair can be clustered in a group. Considering this approach the researcher can have a visual map of the subjects and the clusters help the interpretation of results in a simpler and more understandable manner.

The Weighted Euclidian distance is the Euclidian distance from standardized variables. The Mahalanobis distance differs from the Weighted Euclidian distance because, instead of using the diagonal matrix with variances to standardize the variables, it uses the complete variance and covariance matrix, which means that the relation between the variables are included in the analysis (they are not treated as independent as in the Euclidian distance). Rubin<sup>12</sup> and Cochran and Rubin<sup>13</sup> described the Mahalanobis metric for matching.

The mathematic expression of Mahalanobis' distance is as following:

$$d^2(i, j) = (\vec{x}_i - \vec{x}_j)' \Sigma^{-1} (\vec{x}_i - \vec{x}_j), \quad (3)$$

where  $\vec{x}_i$  is the vector of observed variables for the city with agenda ( $i$ ) and  $\vec{x}_j$  for the city without agenda ( $j$ ).  $\Sigma$  is the sample variance-covariance matrix. In the propensity score analysis, we used logistic regression which standardized the variables according to the variances of the cities with social agendas ( $Y=1$ ). In the same sense for the Mahalanobis' distance, we used  $\Sigma$  based on the cities with a social agenda. An additional argument for doing this is the fact that most of the control candidates will be dismissed, justifying the use of the variability from the sample of cities with social agendas.

The procedure will be similar to that of the propensity score. We used the same randomized list of cities with a social agenda. We selected the first city in the list, and then we found the control candidate with the smallest distance. The control city is taken out from the candidates' sample. The second city with a social agenda is selected and also the pair of control candidates with the smallest distance, and so on. After the whole list of cities with agendas have its pairs, the procedure is performed again, until two control candidates for each city with an agenda are found.

### **Mahalanobis distance matching within propensity score Calipers method**

In an effort to obtain an optimal matching considering both previous methods, we now consider a hybrid system of matching. We used the same randomized list of cities with an agenda. The first city with an agenda is selected and all control candidates with propensity score differences lower than a constant previously chosen are selected (we used 0.2 times the standard deviation of the general propensity score)<sup>10</sup>. Among these control candidates the nearest Mahalanobis metric defines the final control for the matching. If there is no control candidate within the Calipers, the closest propensity score is used to define the final control. The procedure runs until each city with an agenda has one control and then the procedure is performed again to find a second control in the sample without the controls already selected.

### **Statistical comparison among the matched samples**

In order to compare if the matching was satisfactory it was performed a comparison between the cities with and the selected ones without an agenda regarding the matching variables (three comparisons for each variable). For each of the five variables we performed a generalized linear model<sup>14</sup> considering the appropriate distribution. The matching variables were considered as outcome meanwhile the group was considered as co-variable. The final model for PS and VACC considered Gamma distribution because of their asymmetries, and for DHII, DHIE and ALPHA the model considered the Normal distribution. All models were performed with link identity and to account for the matching we specified a correlation structure.

## **RESULTS**

It was found three different matching samples for the 34 cities with a social agenda, one for each method and they did not coincide.

The logistic regression considered for propensity score matching was:

$$\log \left\{ \frac{\pi(x)}{1-\pi(x)} \right\} = -13.105 + 6.66e^{-7}PS + 5.524DHII + 3.183DHIE + 0.045ALPHA + 0.01VACC. \quad (4)$$

This regression presented Hosmer and Lemeshow goodness of fit test<sup>15</sup> of 9.948 with p-value of 0.269, indicating a good fit for this model. We also performed diagnosis analysis (data not shown) and the results were adequate for a non-predictive model.

Results in Table 1 compares groups (with and without social agenda) for four situations: general sample of the entire Paraná state, control group provided by propensity score matching, control group provided by Mahalanobis's distance matching and Mahalanobis with Calipers group selection, respectively. For each matching variable we fixed the group with social agenda as baseline, which means that we compared each of the groups samples with this baseline group. Table 1 shows that the groups differ in DHII and ALPHA for a significance level of 0.05 only for the Mahalanobis' matching and for the total sample (general sample of Paraná state). For the general sample one extra difference was observed for HDIE.

The simplest matching procedure, the propensity, seems to produce a non-exposed group that is similar to the exposed group with respect to the matching variables. The last group, Mahalanobis within Calipers, also presented no differences; however, its PS value is higher (but not significantly) when compared to the exposed group. The sample of non-exposed cities still has differences with relation to the cities with a social agenda when matching was performed by Mahalanobis' method, presenting very similar averages when compared with the general non-exposed sample (probably very similar to a random sample).

Table 2 presents mean differences and standardized difference (%) for each variable of the matching (same statistics used by Rosenbaum and Rubin<sup>10</sup> in their Table 2). Regarding mean difference and also standardized difference the Mahalanobis matching presented the most different results. Regarding mean difference the Propensity matching was better for PS and DHIE whereas Mahalanobis within Calipers was better for DHII and VACC. Finally, regarding standardized differences, Propensity method presented the best result for PS meanwhile Mahalanobis within Calipers presented the best result for all the other variables.

## DISCUSSION

The main purpose of matching is the definition of a group which can be compared with the exposed group without or with reduced confounding and selection bias. We have proposed three methods of control group selection and compared them. Usually in observational studies there is no random selection of control (non-exposed) group, since it could imply in bias because the exposed group was not randomly selected<sup>2</sup>. In that way it is important to consider some method for controlling potential confounding related to bias selection. For example, propensity score can be used to adjust treatment, intervention or exposure effects, through matching, stratification, weighting or including the propensity score as control variable in some statistical models. The matching group has the purpose of making both group's distributions similar, comparable, to guarantee that the association that was estimated is due to the intervention or exposition<sup>11,16</sup>.



Table 1. Imbalance between the variables according to the groups: mean differences test.

Co-variables	Group	n	Mean	Standard Deviation	p-value
PS	With social agenda	34	34668	64223	Baseline
	Without social agenda (general)	363	23067	90408	0.101 <sup>a</sup>
	Without social agenda (propensity matched)	68	34423	68569	0.981 <sup>b</sup>
	Without social agenda (Mahalanobis matched)	68	40482	192208	0.753 <sup>b</sup>
	Without social agenda (Mahalanobis within Calipers)	68	55870	200511	0.406 <sup>b</sup>
HDII	With social agenda	34	0.672	0.048	Baseline
	Without social agenda (general)	363	0.651	0.046	0.009 <sup>c</sup>
	Without social agenda (propensity matched)	68	0.674	0.054	0.626 <sup>d</sup>
	Without social agenda (Mahalanobis matched)	68	0.651	0.045	0.009 <sup>d</sup>
	Without social agenda (Mahalanobis within Calipers)	68	0.672	0.056	0.904 <sup>d</sup>
HDIE	With social agenda	34	0.852	0.059	Baseline
	Without social agenda (general)	363	0.831	0.042	0.009 <sup>c</sup>
	Without social agenda (propensity matched)	68	0.852	0.050	0.906 <sup>d</sup>
	Without social agenda (Mahalanobis matched)	68	0.831	0.041	0.053 <sup>d</sup>
	Without social agenda (Mahalanobis within Calipers)	68	0.852	0.049	0.886 <sup>d</sup>
ALPHA	With social agenda	34	78.14	5.64	Baseline
	Without social agenda (general)	363	76.13	4.14	0.009 <sup>c</sup>
	Without social agenda (propensity matched)	68	77.94	4.50	0.629 <sup>d</sup>
	Without social agenda (Mahalanobis matched)	68	76.07	3.93	0.035 <sup>d</sup>
	Without social agenda (Mahalanobis within Calipers)	68	78.12	4.58	0.962 <sup>d</sup>
VACC	With social agenda	34	83.43	12.26	Baseline
	Without social agenda (general)	363	83.14	10.38	0.874 <sup>a</sup>
	Without social agenda (propensity matched)	68	82.05	8.24	0.583 <sup>b</sup>
	Without social agenda (Mahalanobis matched)	68	82.30	10.15	0.647 <sup>b</sup>
	Without social agenda (Mahalanobis within Calipers)	68	84.12	9.91	0.795 <sup>b</sup>

<sup>a</sup>Generalized linear model With Gamma distribution and link identity;

<sup>b</sup>Generalized linear model With Gamma distribution and link identity for covariate structure;

<sup>c</sup>Generalized linear model With Normal distribution and link identity;

<sup>d</sup>Generalized linear model With Normal distribution and link identity for covariate structure.

Table 2. Means and standardized differences for the variables of matching according to the general sample, control groups by propensity method, Mahalanobis and Mahalanobis within Calipers.

	Mean difference				Standardized difference %			
	General	Propensity method	Mahalanobis' method	Mahalanobis within Calipers	General	Propensity method	Mahalanobis' method	Mahalanobis within Calipers
PP	11602	246	-5814	-21202	14.80	0.31	-7.41	-27.04
IDHR	0.022	-0.002	0.021	0.001	45.670	-3.756	44.575	1.149
IDHE	0.021	-0.001	0.021	-0.001	3.308	-1.059	40.832	-0.106
ALFA	2.01	0.20	2.07	0.02	40.61	4.04	41.89	0.36
VACI	0.29	1.38	1.13	-0.69	2.56	12.19	9.99	-6.05

Note: the standardized difference % is  $100(\bar{x}_{exposed} - \bar{x}_{matched}) / [(\bar{s}_{exposed}^2 + \bar{s}_{reservoir}^2)]^{1/2}$  were  $\bar{x}_{exposed}$  and  $\bar{x}_{matched}$  are sample means in the exposed group and in the matched/control group and  $\bar{s}_{exposed}^2$  and  $\bar{s}_{reservoir}^2$  are the sample variances in the exposed group and in the reservoir control group (all control candidates).

We conducted a selection of two non-exposed cities to be matched with each of the 34 cities with a social agenda in the state of Paraná, Brazil. Furthermore, we used the same data set and we considered the same variables for the three methods of selection. The sample variances and co-variances matrix of the cities with a social agenda were applied to the Mahalanobis' distance method, since the logistic regression analysis had as outcome the social agenda implemented from at least 2004 ( $X = 1$ ), which uses the standardization of the same sample (cities with social agenda).

Three methods for multivariate matched sampling were illustrated, which presented different matching results. The nearest propensity score, that requires less computation effort showed successful reduction of bias in the covariates for the state of Paraná. This method has been considered as a good strategy to improve inferences in studies which are not randomized; however, it should not be used as exclusive method for controlling bias selection<sup>17,18</sup>.

The Mahalanobis' method, which is expected to be very successful in reducing bias, did not present very good results in our case. This method used to be the most usual matching method in studies with several predictors, especially when the predictors present correlation with each other<sup>12,19</sup>. Despite the fact that in the case of Paraná state did not present very good bias reduction (two from five control variables still present differences between groups) it has been considered a useful tool to determine similarities between exposed and non-exposed groups<sup>12,13,19</sup>. In our results Mahalanobis distances did not provide a good match because of the presence of high magnitudes for the distance. Even considering the PS variability in the variance-covariance matrix, the scale of this variable may be too high for the matching, and the others might have low weight in the distance.

Finally, the Rosenbaum e Rubin<sup>10</sup> matching method combining propensity score and Mahalanobis' distance, presented good results even for PS (the biggest variability). The Rubin<sup>12</sup> and Rosenbaum and Rubin<sup>10</sup> percentage reduction in bias were not applied

here because it is unstable when the initial bias is small (for most of the variables in our matching we had a bias lower than 20%). However, Table 2 shows the mean differences after matching.

The two methods which used propensity score presented, in our case, a better result than Mahalanobis's distance by itself, with smaller standardized differences and non-significant difference between two groups after matching. The Mahalanobis's distance within Calipers method is similar to the nearest Mahalanobis' distance method but it adds an additional restriction, only if the control's propensity score is within a certain radius (caliper). Thus, in this method, it is possible that a city with social agenda cannot be matched to a control city. These Calipers can avoid bad matches. Here, when within Calipers area was empty we just ignored the caliper which provided a worse result but, on the other hand, provided matching for the whole sample.

Since, the current study has examined only three methods in one sample, further empirical studies of multivariate matching methods are required. Moreover, it would be interesting to know when Mahalanobis' distance will be a better match than propensity score. It is possible that some special variance-covariance methods could be proposed for matching variables that are not normally distributed. It was not the case, but if it was considered a categorical variable for the matching, the Mahalanobis' distance could not be applied.

The use of propensity Calipers guarantees that the difference in propensity score cannot be high (the cut-off is pre-defined). Even here that the exclusion of Caliper was allowed, comparing all results in Table 2, this method performed better than propensity by itself for most of the variables, with the only exception of the variable PS (the highest variance). If we did not allow the exclusion of Caliper or if we have used another cut-off for the Mahalanobis methods, this method would present even better results for the matching. However, in this method, many exposed cities would be with no controls because the distances were not very low.

The number of scientific articles applying propensity score is not very high, but it is increasing rapidly. The main reason is the effective results in controlling for potential confounding in comparison to usual multiple regressions<sup>18</sup>. In Epidemiology it is considered an efficient method for selection bias correction but is still not very often used. Its use for reducing bias selection may be efficient due to the fact that the propensity between the final two groups samples are very similar which means their distribution regarding predictive variables are similar, allowing groups comparison<sup>4,16,17</sup>.

In conclusion, the propensity, which is a simple linear score, presented an effective result in providing a non-exposed sample which showed better results than the general sample to be compared with the sample with social agendas. However, the Mahalanobis within Calipers presented even better results and it is recommended to be used. Finally it is important to note that in this specific project the selection of a sample to be compared to the exposed sample is extremely necessary, since the second phase of the study had the main purpose of verifying the real non-agenda exposition to make sure the comparison of phase 3 (comparison of the groups regarding MDG) would bring the information we were looking for.

## ACKNOWLEDGMENTS

We would like to thank The National Council for Scientific and Technological Development (CNPq) support (process n° 409821/2006-3 - Auxílio a pesquisa -Ed 262006 Determ Soc-Edital M (MCT/CNPQ/MS-SCTE-DECIT) and CNPq (processo n° 142438/2008-1) support for the congress *XI School of regression models* (“XI Escola de Modelos de Regressão”) in Recife (Brazil).

## REFERENCES

1. Rosenbaum PR. Optimal matching for observational studies. *Journal of the American Statistical Association* 1989; 84(408):1024-32.
2. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed Philadelphia: Lippincott-Raven; 1998.
3. Miettinen O. Estimability and estimation in case-referent studies *Am J Epidemiol* 1976 Feb; 103(2): 226-35.
4. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): 41-55.
5. Pereira MG. *Epidemiologia: Teoria e prática*. Rio de Janeiro: Guanabara Koogan; 2000.
6. Westphal MF (coord.). *Relatório Final: Saúde e Desenvolvimento Local: análise dos progressos em relação aos Objetivos de Desenvolvimento do Milênio relacionados à saúde, nas cidades brasileiras que desenvolvem agendas sociais [relatório de pesquisa]*. São Paulo: Faculdade de Saúde Pública; 2009.
7. Cerqueira F, Facchina M. Agenda 21 e os objetivos de desenvolvimento do milênio: oportunidades para o nível local. *Cadernos de Debate* N° 7. Agenda 21 e Sustentabilidade. Ministério do Meio Ambiente. Secretaria de Políticas para o Desenvolvimento Sustentável. Brasília: Ministério do Meio Ambiente; 2005.
8. Westphal MF. O movimento cidades/municípios saudáveis: um compromisso com a qualidade de vida. *Ciênc Saúde Coletiva* 2000; 5(1): 39-51.
9. Akerman M. *Saúde e desenvolvimento local: princípios, práticas e cooperação técnica*. São Paulo: Hucitec; 2005.
10. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 1985; 39(1): 33-8.
11. D'Agostino RBJ. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; 17(19): 2265-81.
12. Rubin DB. Bias Reduction using Mahalanobis-metric matching. *Biometrics* 1980; 36: 293-98.
13. Cochran WG, Rubin DB. Controlling bias in observational studies: A review. *Sankhya: The Indian Journal of Statistics, Series A* 1973; 35: 417-46.
14. Nelder JA, Wedderburn WM. Generalized linear models. *J R Statistic Soc A* 1972; 135 (3): 370-84
15. Lemeshow S, Hosmer Jr DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982; 115(1): 92-106.
16. Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Serv Outcomes Res Methodol* 2001; 2(3-4): 169-88.
17. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol* 2006; 59(5): 437-47.
18. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiol Drug Saf* 2004 Dec; 13(12): 841-53.
19. Mahalanobis PC. On the generalized distance in statistics. In: *Proceedings of the National Institute of Sciences of India* 1936; 2(1): 49-55.

Received on: 05/15/2013

Final version presented on: 11/19/2013

Accepted on: 03/17/2014