Journal of Statistics on Health Decision

ORIGINAL ARTICLE

Evaluation of cardiovascular disease risk factors with Propensity Score Matching and Coarsened Exact Matching: Nepalese post-seismic observational data.

Tiago Adrega[1], Anabela Rocha[2], M. Cristina Miranda[2,3]

[1]Serviço de Cardiologia, Centro Hospitalar Baixo Vouga, Aveiro
[2]ISCA e CIDMA, Universidade de Aveiro
[3]CEAUL, Universidade de Lisboa

ABSTRACT

With observational data, an important step of the research process is skipped, resulting in some restrictions to make inferences concerning the treatment effects. Some methodologies have been developed in order to reduce the imbalanced in the samples of treated and control units. Propensity Score Matching (PSM) is still one of the most common approaches applied but Coarsened Exact Matching (CEM) appears to produce better results, most of the times in which it is used. This work illustrates the application of each of the two techniques to a set of data from the Nepal population. Our aim is to compare the two methodologies and evaluate in what way their use adds information about the prevalence of Cardio Vascular Disease (CVD) risk. Data refers to a remote village population that was separated in two groups after the incidents of the May 2015 earthquake. The study was carried out during a humanitarian mission in Nepal, aimed to provide medical care to the people of Sindhupalchok, a northern Nepalese region, with approximately 1200 inhabitants. With the seismic event this population got separated in two groups of dislodged individuals: victims that stayed nearby the village area and those who went towards Kathmandu looking for support in temporary settlements. Both these populations were supported by the medical mission. Cross-sectional data was collected approximately 18 months after the earthquake and included demographic data, anthropometric data, previous medical history, CVD risk factors and health behaviors. The assessment of CVD risk factors and health behaviours was based in a question-by-question guide provided by the WHO.

In order to compare both approaches we computed two imbalance measures, L1 and Percent Bias Reduction (PBR). The results show that CEM dominates PSM. From the application of the two approaches we find that the results are generally in agreement but CEM methodology allowed to highlight some data features not seen before with PSM.

Introduction

In order to evaluate some particular treatment effect, if and when it is possible, a designed random treatment allocation ensures the existence of two groups with similar baseline characteristics such that treatment status will not be confounded by any different values of covariates presence so it is possible to estimate the effect of the treatment comparing the two groups. Unfortunately, most of the times that is not a feasible scenario, particularly in social and clinical sciences. This motivates the use of observational studies which, as Cochran [1] points out, has allowed to show some important evidences in human health despite not controlling for the conditions of a random experience. Often being the only possible way of conducting a research project, it is of the most importance to overcome eventual confounding covariates effects in such empiric investigations so the conclusions may reproduce real consequences of the effect of some treatment on the subjects.

When dealing with observational data, an important step of the research process is skipped, resulting in some restrictions to make inferences. Confounding variables might appear, misleading results about eventual existing causal effects. As Rosenbaum notes [2], in any scientific experiment, the experimenter controls the process of assigning the treatments to the subjects under study, in such a way that she can guarantee to have two groups with similar characteristics receiving different treatments so they may be comparable. In an observational study, it is not possible to control that process. The reasons for this lack of control may be diverse due to restrictions of different sources.

Most of the research studies in the areas of health and clinical science result from empiric investigations, where there is not the possibility of randomly assign subjects to different treatments or procedures.

If no action is taken, the most probable existence of confounding variables will turn the study to produce erroneous results. It is desirable to reduce or, better, to eliminate the effects of confounding variables when using observational data to properly estimate treatment effects.

Keywords:
Cardiovascular disease risk factors; Propensity score matching; Health behaviours; Coarsened exact matching; Observational studies.

Corresponding author:
M. Cristina Miranda
cristina.miranda@ua.pt

Conflict of interest:
The authors declare no conflict of interests

Received date: 06.APR.2021
Revised date: 25.JUN.2021
Accepted date: 09.NOV.2021
First published: 14.DEC.2021

Different methods are proposed in the literature to approximate those data to a typical set of data from a designed experiment. Some authors propose analytical adjustments in order to reduce the existing bias like matching and stratification [2][3]. Causal inference might be considered as a relatively new branch of statistics but, after the initial lift-off due to the studies of Cochran [1] and the reference text of Rubin and Rosenbaum and [2], there has been a well-developed literature with different proposed causal inference techniques. Two of the most frequently used of those techniques are Propensity Score Matching (PSM) [4,5] and Coarsened Exact Matching (CEM) [6,7]. PSM is one of the first approaches to be proposed and has been widely used. CEM appeared more recently and has proven to produce better results, resulting in estimates of causal effects that are less biased and with lower variance, regardless the dimension of the sample [8].

The goal of the present study is to apply the two methodologies and compare the results within an evaluation of the prevalence of Cardio Vascular Disease (CVD) risk factors in two cohorts of a native Nepalese population. The April and May 2015 earthquakes imposed critical social and epidemiological pressures, resulting in critical lifestyle changes, namely regarding CVD risk factors [9]. Various organizations provided support to these people and engaged in various activities aiming to opportunistically improve this CVD burden. This study focus in the two groups of people arising from a common origin, a remote village in Nepal that was affected by the May 2015 earthquake. People were separated after the incidents, a group remained in the village and the others went to a camp from one of the aid organizations, where better facilities could be provided. We aim to understand if access to camp facilities has an effect on the prevalence of heart diseases. Non-communicable diseases are the leading causes of death in developed and developing countries worldwide [10]. Nepal is an example of this paradigm - from 2005 to 2015, the ischemic heart disease increased around 25.3% and brain vascular disease increased 25.7% [11].

The paper is organized in sections as follows: after an introduction we present the methods applied, PSM and CEM and we present the measures of balance used to compare both of them. In the third section we briefly describe the data and in the fourth section we present the results. The paper ends with some conclusions.

Methods

Propensity Score Matching (PSM)

Suppose we have units available with observed values of some vector of covariates $\mathbf{x}$. Let $\mathbf{x}_{[i]}$ be the covariate for the i element and $T_{[i]}$ the treatment assignment for the same unit, being

$$T_{[i]} = \begin{cases} 1 \\ 0 \end{cases} \text{ if unit } i \text{ recives treatment, otherwise.}$$

Let $\pi_{[i]} = P\left(T\left[i\right] = 1\right)$ so $1 - \pi_{[i]} = P\left(T\left[i\right] = 0\right)$, $\left(0 < \pi_{[i]} < 1\right)$ and assume that the assignment for different i, l units is independent. In this case,

$$P\left(T_{[1]} = t_1, ..., T_{[n]} = t_n\right) = \prod_{i=1}^{n} \pi_{[i]}^{t_i} \left[1 - \pi_{[i]}\right]^{1-t_i}$$

In any randomized experience, this probability is known. This is not the case in an observational study where the $\pi_{[i]}\, S$ are unknown. However, if we have reasons to believe that the $\pi_{[i]}\, S$ only depend on the observed covariates $\mathbf{x}_{[i]}$, then there is a function λ such as $\pi_{[i]} = \lambda\left(\mathbf{x}_{[i]}\right), i = 1, \cdots, n$. This means that any two units with the same values of $\mathbf{x}$ have the same probability of being assignment to receive the treatment. An observational study with this property is said to be *free of hidden bias* and the λ function is called *propensity score* [2]. The previous probability then becomes

$$P\left(T_{[1]} = t_1, \cdots, T_{[n]} = t_n\right) = \prod_{i=1}^{n} \lambda\left(\mathbf{x}_{[\mathbf{i}]}\right)^{t_i} \left[1 - \lambda\left(\mathbf{x}_{[\mathbf{i}]}\right)\right]^{1-t_i}$$

One simple approach to adjust for the bias is to stratify on the covariates $\mathbf{x}$. Some units from the total of n units are chosen and reorganised into S non overlapping strata with n_s units falling in stratum s. The units are renumbered in this mode so the lth unit in stratum s has treatment assignment T_{sl} and covariate $\mathbf{x}_{sl}$ If $\mathbf{T} = \left(T_{11}, \cdots, T_{Sn_S}\right)^T$ and m_T^S is the number of treated units in stratum s, the $m_T^S = \Sigma_l T_{sl}$ and $\mathbf{m} = \left(m_T^1, ..., m_T^S\right)^T$.

Exact stratification

In an exact stratification, each unit with the same value of $\mathbf{x}$ belongs to the same stratum, *i.e.*, $\forall l, j, s, \mathbf{x}_{sl} = \mathbf{x}_{sj}$. Exact stratification is difficult to achieve in most cases. This happens because it becomes harder to find units with the same values of $\mathbf{x}$ when we have a high number of covariates or if they do not assume discrete values.

Matching on $\boldsymbol{x}$

Matching a sample on $\mathbf{x}$ corresponds to define some criteria on S, $\mathbf{m}$ and $\mathbf{n}$ and then making a stratification that meets such criteria. An exact matching, like an exact stratification, is often impossible to obtain due to the same reasons.

Propensity score (PS)

When there are many covariates, the most realistic scenario is to find many units with unique values of $\mathbf{x}$ so exact matching or stratifying is not possible, meaning that is not possible to find units that are homogeneous with respect to $\mathbf{x}$. Still, it may be feasible to find sets that although not homogeneous on $\mathbf{x}$ show similar distributions of $\mathbf{x}$. This means we have covariate balanced study.

The propensity score is the conditional probability of receiving the treatment given the observed covariates $\mathbf{X}$. Each subject l from stratum s, if randomly selected, has probability $1/n_s$ of being chosen from stratum s; then it has probability λ of being selected to receive treatment, so λ is the marginal probability of a unit in stratum s receiving the treatment.

$$\lambda = \frac{\sum_{l=1}^{n_s} \pi_{sl}}{n_s}$$

If there is no hidden bias, it only requires homogeneous sets of data on $\lambda(\mathbf{X})$ (rather than $\mathbf{X}$) to form matched or stratified balanced sets of data.

Rosenbaum and Rubin, [5] proved the balancing property of the propensity score.

Theorem:

If $\lambda(\mathbf{X}) = \Lambda$, then

$$P[\mathbf{X} = \mathbf{x}_s \mid \lambda(\mathbf{X}) = \Lambda, T = 1] = P[\mathbf{X} = \mathbf{x}_s \mid \lambda(\mathbf{X}) = \Lambda, T = 0].$$

This means that if we pick some value for the propensity score (PS), Λ, and randomly choose one subject among all which have that value of PS, then for that subject, treatment assignment T is independent from the covariate value given the value $\lambda(\mathbf{X}) = \Lambda$.

So, conditional on the propensity score value, the distribution of observed covariates will show similar characteristics between treated and untreated subjects. So, treatment assignment and response are conditionally independent, given $\mathbf{X}$ and it is assumed a common support between the treatment and control groups, *i.e.* a common region for distribution values.

Austin [12] identifies four different methods to apply PSM: matching on the propensity score, stratification on the propensity score, inverse probability of treatment weighting using the propensity score, and covariate adjustment using the propensity score. He also examines different ways of estimating the unknown PS. Besides the most frequent approach used to compute this probability - the logistic regression model - he refers the use of machine learning [13] and neural networks [14], among others.

In this work, the unknown probability of each individual in a sample being assigned to a treatment is estimated from the data using a logistic regression model: treatment assignment is regressed on the set of observed covariates. The propensity score then allows matching of individuals in the control and treatment conditions with the same likelihood of receiving treatment. Thus, a pair of participants (one in the treatment, one in the control group) sharing a similar propensity score are seen as equal, even though they may differ on the specific values of the covariates.

Once propensity scores have been calculated, we use some criteria (in this case we have chosen the nearest neighbour procedure which consists in the minimum absolute difference of the two scores) to find individuals in the control group that will have similar propensity scores to those in the treatment group [15].

Imbems, [16] describes two different approaches to perform the matching process after computing the values of PS by through a logit model. In the first one, suppose we have a number of treated units smaller than the number of control units. The treated units are reordered by its PS estimated value and, beginning with the largest PS estimated value, it is matched with the one in the control group to which corresponds the closest estimated PS value. So, if the treatment group has units, the resulting matched subset sample has units. The second approach follows the work of Imbems et al [17] and focus on the average treatment effects for subsets of the covariate space as a way to guarantee substantial overlap for the covariate distributions in the two groups. To achieve this, extreme observations of PS estimated values are dropped. The resulting matched sample is a trimmed sample pruned of the most extreme values that would be difficult to match between the two groups. Other different matching algorithm to match units from the two sets based on the PS are proposed in the literature [18]. All of these seek to improve the trade off between the balance of the groups of treatment and control and the size of the samples.

Caliendo and Kopeneig [18] summarize the necessary steps to apply PSM: start to estimate propensity scores, choose a matching algorithm, check overlap/common support, estimate quality/effect and, finally, sensitive analysis, with respect to eventual unobserved heterogeneity or failure of the common support condition.

Coarsened Exact Matching (CEM)

Coarsened Exact Matching (CEM) or "Cochran Exact Matching" (recognizing Cochran merits in the primordial study of observational data with the first sub-classification-based method [19]) is a matching method, proposed in [20] that belongs to the Monotonic Imbalance Bounding (MIB) class of matching methods for causal inference, introduced in the same paper. CEM applies exact matching after each variable have been separately coarsened. MIB classes were developed to control and avoid the increase of the global imbalance on the variables; without requiring extra assumptions on the data, allows to reduce the imbalance of one variable without affecting the maximum imbalance for the others. Although a MIB method can not guarantee simultaneously a bound on the level of imbalance and a specified number of matched observations, it allows the user to choose the maximal imbalance ex ante and produces a matched sample size ex post. This is an advantage for observational data analyse, since data is not under control of the investigator, and so reducing bias rather than inefficiency is the main focus [20].

For a sample of n units or elements, let T_i be the treatment variable for the unit i, where $T_i = 1$ if i unit received treatment (and belongs to the "treated" group) and $T_i = 0$ if unit i did not receive treatment (and belongs to the "control" group). The outcome variable is Y, where $Y_i\,(0)$ is the "potential outcome" for observation i if unit i does not receive treatment and $Y_i\,(1)$ is the potential outcome if the unit i receives treatment. For each unit i, only one potential outcome is observed. This can be expressed by the condition $Y_i = T_i Y_i(1) + (1 - T_i)Y_i(0)$ indicating that $Y_i\,(0)$ is unobserved if unit i belongs to the "treated" group and that $Y_i\,(1)$ is unobserved if unit i belongs to the "control" group. Without loss of generality, we assume that when we refer to unit i, we assume it as a treated unit, $Y_i\,(1)$ is observed while $Y_i\,(0)$ is unobserved and thus must be estimated by matching it with one or more units from a given set of the control units. Let $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_k)$ be a data set, where each $\mathbf{x}_j$ is a column vector of the observed values of pretreatment variable j for the observations, *i.e.*, $\mathbf{X} = (x_{ij}), i = 1, ..., n; j = 1, ..., k$. We note by T the set of indexes for the treated units, $n_T = \#T$ the number of treated units and, in a similar way, by C and $n_C = \#C$ for the control units, with $n_T + n_C = n$.

According to the proposing authors, *CEM* requires three steps: (1) Coarsen each of the original variables in $\mathbf{X}$, resulting in $C(\mathbf{X})$. (2) Apply exact matching to $C(\mathbf{X})$, which consists in sorting the observations into, say, strata $s \in S$. (3) Strata containing only control units are discarded; strata with treated and control units are retained; and strata with only treated units are used with extrapolated values of the control units or discarded.

Let T^s be the treated units in stratum s and $m_T^s = \#T^s$, and, in a similar way for the control units, C^s and $m_C^s = \#C^s$. The number of matched units are, respectively for treated and control units, $m_T = \Sigma_{s \in S} m_T^s$ and $m_C = \Sigma_{s \in S} m_C^s$. For each matched unit i in stratum s associate a CEM-weight defined by $w_i = 1$, if $i \in T^s$, $w_i = \frac{m_C}{m_T}$, if $i \in C^s$, and $w_i = 0$, if i is unmatched unit. These weights can be used to calculate the sample average treatment effect on the treated, $SATT = \frac{1}{n_T} \Sigma_{i \in T} [Y_i(1) - Y_i(0)]$ as described in [20].

The Coarsening Exact Matching Approach has been used in several health research studies as [21], [22], [23] and [8]. For practical applications, the R program can be used whith cem function which is available in *MatchIt* [6].

Imbalance measures

The main goal of matching is to improve balance and this is also the way to measure its success jointly with the number of observations that are kept for the analysis [3]. Different measures of unbalanced are used to assess for the degree of balance achieved. Comparing means and standard deviations of each covariate in both groups, comparing the histograms or, as Ho et al [3] prefer, plotting and compare the distributions by means of a quantile-quantile plot (QQ-plot). Empirical cumulative density function values should be close or, equivalently, the differences values between the treated group and the control group should be close to zero. Two different measures of evaluating unbalancing and heterogeneity are the L_1 metric and the Wald test [8]. The L_1 measure in a non-parametric context and the Wald test when distributional assumptions may be assumed. These measures provide a quantification of the extent to which treatment and control groups distributions differ.

Pan and Bai [24] identified graphical and statistical ways to evaluate the covariate balance for the PSM methodology. One of the referred measures is the standardized bias () defined in the work of Rosenbaum and Rubin [25] as

$$SB = \frac{M_1(\mathbf{x}) - M_0(\mathbf{x})}{\sqrt{\frac{V_1(\mathbf{x}) - V_0(\mathbf{x})}{2}}} \times 100\%,$$

where $M_1(\mathbf{x})$ and $V_1(\mathbf{x})$ are the values of the mean and variance for the covariate units in the treatment group, respectively, and $M_0(\mathbf{x})$ and $V_0(\mathbf{x})$ are the values of the mean and variance for the covariate units in the control group, respectively. Values of standardized mean differences close to zero indicate good balance. Caliendo and Kopeinig [18] consider a value of this measure after matching bellow as signal of success in the process of balancing covariate distributions. One different metric that is also often used is the Mahalanobis distance between each unit in the treatment group and the closest unit in control group, averaged over all units. The percent bias reduction (PBR) [26] is another way to evaluate the quality of the matching process. The bias associated to a covariate $\mathbf{x}$ is $B = M_1(\mathbf{x}) - M_0(\mathbf{x})$; the bias after matching, $B_a = M_{1_a}(\mathbf{x}) - M_{0_a}(\mathbf{x})$ is compared with the bias before matching, $B_b = M_{1_b}(\mathbf{x}) - M_{0_b}(\mathbf{x})$ to obtain PBR:

$$PBR = \frac{B_b - B_a}{B_b} \times 100\%.$$

Empiric results indicate that a value of $PBR \geq 80\%$ represents a reasonable bias percent reduction.

L_1 was proposed in [20] as a measure to access the imbalance between the treated and control groups and and was used to confirm that CEM outperforms other methods for different data sets. Consider $H(\mathbf{x}_j)$ denoting the set of intervals into which the support of variable X_j has been cut and the multidimensional histogram, defined by the Cartesian product $H(\mathbf{x}_j) \times ... \times H(\mathbf{x}_k) = H(\mathbf{X}) = H$. Whereas in the original data the weights are equal to one for all observations in the sample, when a matching method is applied, the observation weights can be calculated based on the relative empirical frequency distributions for the treated and control units, f and g. Let $f_{l_1 \ldots l_k}$ and $g_{l_1 \ldots l_k}$ be the relative frequencies for observations belonging to the cell with coordinates $l_1 \ldots l_k$ of the multivariate cross-tabulation. The L_1 measure is defined by

$$L_1(f, g, H) = L_1(H) = \frac{1}{2} \sum_{l_1 \ldots l_k \in H} |f_{l_1 \ldots l_k} - g_{l_1 \ldots l_k}|,$$

where the notation $L_1(H)$ refers to its dependence on the choice of multidimensional histogram H. Several situations studied by the authors shown that the choice of H is not relevant in the value of the imbalance measure L_1.

Measurement L_1 is based on relative frequencies, which allows comparison of samples with different sizes for the treated and control groups. The measure L_1 takes values in the interval $[0, 1]$ and has an intuitive interpretation: $L_1 = 1$ indicates that the two empirical distributions are completely separate; $L_1 = 0$ indicates global balance, *i.e.*, the distributions exactly match; $0 < L_1 < 1$ indicates the amount of difference between frequencies of the two groups. According to [20], the following interpretation can be made: for example, if $L_1 = 0,6$, then there is a 60% separation between the two distributions, *i.e.*, there is a 40% similarity between the two distributions. Let f_m and g_m be the distributions of the matched treated and control units, corresponding to the distributions f and g of the original data. A good matching method must verify that $L_1(f_m, g_m) \leq L_1(f, g)$.

Statistic analysis

We started the statistical analysis with a descriptive and exploratory study in which we obtained graphs and tables with statistical values that allowed to characterize the marginal and joint distributions of the variables. We then apply PSM and CEM methodologies to obtain balanced samples of the two groups formed accordingly to the place in which they were allocated after the earthquake. Assuming we succeed in reducing imbalance of the two groups, we could then applied statistical inference techniques to assess the impact of the population's resettlement on CVD risks. We ran parametric (T-test and Binomial test) and non-parametric tests (Wilcoxon) to compare some variables related to CVD risk factors and health behaviours for both groups. We also fitted some regression models (logistic model and multiple linear regression model) to estimate some of the identified CVD risk factors and identify behaviours and to identify behaviours and factors with significant influence on these risks. All the decisions were taken comparing p-values with a significance level of 0.05.

Data

This study is an observational, cross-sectional study that was carried out during a humanitarian mission in Nepal, aimed to provide medical care to the people of Sindhupalchok, a northern Nepalese region, with approximately 1200 inhabitants. With the seismic event this population got separated in two groups of dislodged individuals: victims that stayed nearby the village area and those who went towards Kathmandu, looking for support in temporary settlements. Both these populations were supported by the medical mission.

Cross-sectional data was collected approximately 18 months after the earthquake and included demographic data, anthropometric data, previous medical history, Cardiac Vascular Disease (CVD) risk factors and health behaviours. The assessment of CVD risk factors and health behaviours was based in a question-by-question guide provided by the WHO [27–29]. The obtained sample is thus a set of two groups formed accordingly the place they were allocated after the earthquake and the following events. This means that the selection of the elements of each group was not carried out respecting the principle of randomness, as it should be for the purpose of statistical inference. This is often the case in this type of studies due to reasons such as ethics, practical issues or even non-viability [30]. Analysis of the resulting data can thus lead to biased conclusions.

The original data is a sample of 230 people separated in two groups, depending on the location. A group of 143 people remained at the village and 87 were logged at the camp. A set of records were taken considering the goal of the study. Some indicators were registered like: Body Mass Index, (*BMI*), Hypertension (*CVR_HTN*), Alcohol (*CVR_Alcohol*), Smoking (*CVR_Smoking*), Diabetes (diabetes mellitus) (*CVR_DM*) and Overweight (*CVR_Overweight*), as dependent variables of interest. The aim is to compare the values of each one of these risk factors between the two groups. Does reallocation affect the behaviours of these variables?

To evaluate the existence of significant differences of these variables among the two groups it is fundamental to have similar characteristics with respect to the covariates considered like , , and so we can minimize the effect of confounding factors. If we had a randomized experience this would not be necessary. However, data shows different distributions on those covariates conditional to the origin factor. This can be seen in Figures 1 and 2, where it is visible that older people, more women and less educated ones remained at the village.

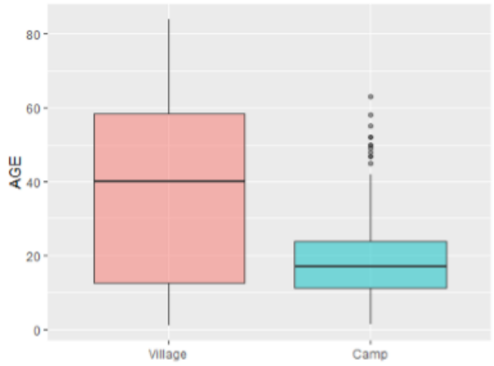

FIGURE 1 AGE distributions (original data).

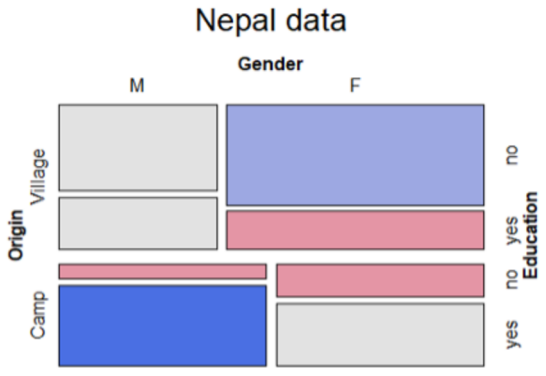

FIGURE 2 Gender, Education and Origin cross original data.

Results

All the results were obtained with R language [31]. The package *Matchit* [32] was used for PSM results whereas cem [6] package was used to obtain results with the *CEM* methodology.

Results with PSM

With PSM we were able to obtain two balanced samples relative to a set of chosen covariates: one for the people that stayed nearby the village and another with the individuals in temporary settlements.

This process was applied to the chosen dependent variables, namely, *BMI*, *CVR_HTR*, *CVR_Alcohol*, *CVR_Smoking*, *CVR_DM* and *CVR_Overweight* with the covariates *age*, *gender*, *education* and *exercise*.

Is the data balanced? A graphical analysis suggests they are not. The Figure 3 illustrates the results of the matching process with nearest neighbour matching based on the PS estimated with the logit model.

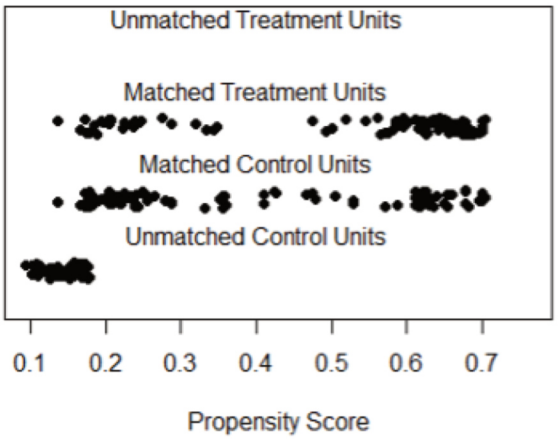

FIGURE 3 Distribution of propensity scores

Distribution of propensity scores

As we can see in Figure 3, the new matched groups are much more similar with respect to the co-variables. This way the comparisons of the dependent variables between the two groups (in village and reallocated in temporary settlements) give more realistic and accurate conclusions.

With the matched samples we proceed to analyse the data. For the body mass index, (*BMI*) we assumed asymptotic Normality and, having not detected significant differences between variances of the two groups (Levene's test $p - value = 0.42$), we compared the mean values with a one sided t-test, concluding that this index tends to be higher for the nearby village located people ($p - value = 0.01$).

As the hypertension variable is a dichotomous one, we performed a proportion test, concluding that the proportion of people with hypertension is significantly greater in the nearby village group then for the ones that went to the camp ($p - value = 0.04$).

We also noticed more severed smoking and alcohol habits in the same group. In these cases, a non-parametric test revealed a significant greater median value for both variables (Wilcoxon's test $p - value = 0.001$ for alcohol and $p - value = 0.04$ for smoking).

Differences in the probability of having *HTN* risk for the two groups are suggested by the box plots in Figure 4, where there seems to be a higher *HTN* risk for older people and with greater BMI.

A logistic regression model with *HTN* as dependent variable *Age* and *BMI*, and *Origin* as explanatory variables is statistically significant ($p - value = 0$ for the deviance reduction test). We may conclude that for people with the same *BMI* and *Origin*, it is expected that the odds of having *HTN* risk increases 1.09 for each year older. Also, for people with the same *Age* and *Origin*, it is expected that the odds of having *HTN* risk increases 1.13 for each unit more in *BMI* measure. Comparing the *HTN* risk for the two groups, for example, for a 50 years adult with 30 units of *BMI* that remained in the village, the probability of having *HTN* risk is greater (0.66) than for the ones that were reallocated (0.58).

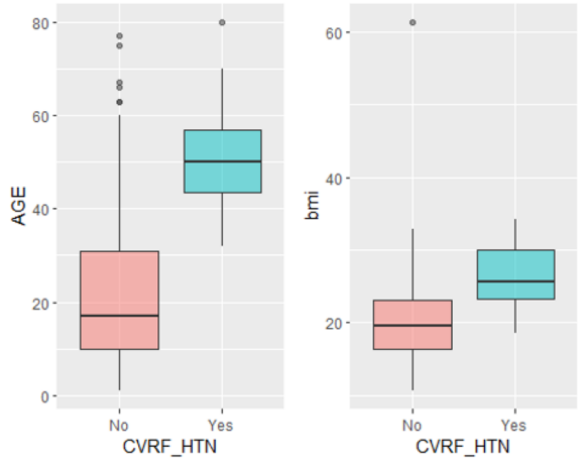

FIGURE 4 Distribution of *AGE* and *BMI*, depending on the *CVRF_HTN* risk

The correlation matrix motivates the study of effects of *Age* and *Education* on *BMI* values (Figure 5).

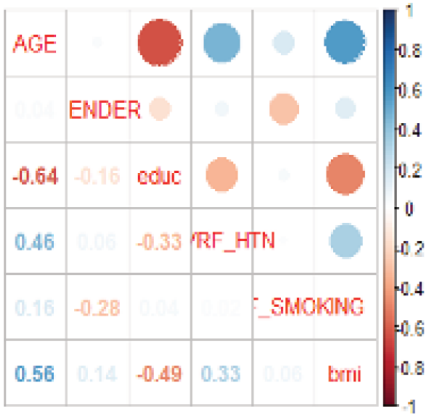

FIGURE 5 Correlation matrix plot

We found a significant multiple regression model ($p - value = 0$) with an estimated equation,

$$\widehat{BMI} = 18.48 + 0.14AGE - 2.09EDUC$$

to explain the BMI variation in terms of age and levels of education. This model explains 31% of *BMI* mean variation. Both variables show significant effects on *BMI* variation. For each year age plus in people with the same education there is an expected increase of 0.14 points in *BMI*; one level more in education of people with the same age gives an expected decrease of 2.09 in *BMI*.

The performed tests reveal significant differences between the two groups. Namely, there is a tendency for greater values of the body mass index, hypertension alcohol and smoking in those people that stayed nearby the village. As far as the exercise, presence of diabetes disease and overweight, we did not establish significant differences among the two groups. The odds of having *HTN* risk increase with *Age* and *BMI*. We can even state that people that stayed in the village have greater probability of *HTN* risk compared with the ones that were logged in temporary settlements. It was also possible to conclude that age and level of education have a significant and positive effect in *BMI* variation – high values of *BMI* are expected for older people with high levels of education. Long-time exposition to the daily living in a provisional camp aimed to provide support to dislodged Nepalese people after the earthquake, might have an effect in some health behaviours and prevalence of *CVD* risk factors.

Results with CEM

We begin by comparing the balanced state of the sample after applying *CEM* methodology. The overall measures of imbalance are computed before ($L1 = 0.579$) and after applying CEM ($L1 = 0.089$). The values of denote a great reduction and there is also an increased percent of the bias reduction, from $PBR = 46\%$ to $PBR = 81\%$

These results make possible to conduct statistic analysis. Applied t-test reveals that the index tends to be higher for the village people than for the realocated group. After a Levène test, not rejecting variances equality ($p - value = 0.61$), we had a $p - value = 0.004$ for the unilateral t-test. As for the hypertension variable, a proportion test was performed, concluding that there is no significant differences in the proportions of people with hypertension in both groups ($p - value = 0.592$). Smoking, alcohol habits and overweight risks were also investigated factors. For all of these risk factors, non-parametric tests indicated significant greater values of medians in the village group, compared to the ones that were in temporary settlements (all $p - values < 0.05$). As far as the exercise and diabetes risks, no significant differences between the two groups were found (all $p - values > 0.05$). A logistic regression model was fitted to the data. The hypertension risk significantly depends on *AGE* and *BMI* leading to the similar conclusions as the ones obtained with *PSM* methodology. Other explanatory variables were considered like *Origin*, *Gender*, *Education* and $CVRF_smoking$ but they didn't present significant effects. *BMI* variation was another matter of interest. A linear regression model was fitted including the *AGE* and *Education* level as significant explanatory variables. The explanatory variables also considered, *Origin*, *Gender* and *Exercise* were not significant for the model. The estimated equation is:

$$\widehat{BMI} = 19.29 + 0.13AGE - 2.70EDUC.$$

Discussion and conlusions

The present study permits to illustrate the proven results of King and Nielsen [33] and Iacus et al. [20] with the Nepal observational data set. With this data we also could observe the dominance of the *CEM* process, when compared to *PSM*, resulting in an improved balanced between the groups of treatment and control, thus increasing the accuracy of the posterior analysis performed.

Table 1 - Imbalance values measures.

Data	L1	PBR
Raw	0.58	0.46
PSM	0.50	0.52
CEM	0.09	0.81

The difference between the two starting point sets of data produced by the two different matching processes, the PSM and the CEM is reflected on the imbalance measures. Comparing with the raw data (before any type of matching), CEM dominates PSM. In Table 1, note the value of L1, closest to zero for the CEM process, reflecting a strong reduction of the imbalance. The remaining indicator measures have higher values also confirming the advantages of using CEM as the chosen matching process.

Application of CEM methodology allowed to highlight some data features not seen before with PSM. Significant differences between the two groups were identified: a tendency for greater values of the body mass index, hypertension, alcohol, smoking and overweight (this risk wasn't even identified as significant when PSM was applied) in those people that stayed nearby the village. The exercise and the presence of diabetes disease did not reveal to have significant differences between the two groups. A significant model was fitted to explain part of the average hypertension risk with age and BMI variation. Also, BMI average variation could be significantly explained by age and education level. The *CEM* methodology allowed to identify overweight as a statiscally different CVD risk factor between the two groups. Behaviour, nutritional, social and physical factors might play an influence, opening the role of the NGO's as potential bridges for consistent health and lifestyle interventions.

Acknowledgements

The authors would like to express appreciation for the support of Associaçao Obrigado Portugal – the Portuguese NGO that runs Camp Hope, a temporary settlement that helped more than 350 people and that still supports the academic life of more than 80 children - and the local partner, Dwarika's Foundation.

This work is supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020.

References

1. Cochran WG, Chambers SP. The Planning of Observational Studies of Human Populations. Journal of the Royal Statistical Society Series A (General). 1965;128:234. https://doi.org/10.2307/2344179
2. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika. 1983;70:41. https://doi.org/10.1093/biomet/70.1.41
3. Ho DE, Imai K, King G, et al. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis. 2007;15:199–236. https://doi.org/10.1093/pan/mpl013
4. He H, Hu J, He J. Overview of Propensity Score Methods. In: He H., Wu P. CD(, editor. Statistical causal inferences and their applications in public health research. Springer, Cham; 2016. pp. 29–48. https://doi.org/10.1007/978-3-319-41259-7_2
5. ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70:41–55. https://doi.org/10.1093/biomet/70.1.41
6. Iacus SM, King G, Porro G. cem : Software for Coarsened Exact Matching. Journal of Statistical Software. 2009;30:1–27. https://doi.org/10.18637/jss.v030.i09
7. Iacus SM, King G, Porro G. Causal inference without balance checking: Coarsened exact matching. Political Analysis. 2012;20:1–24. https://doi.org/10.1093/pan/mpr013
8. Wells AR, Hamar B, Bradley C, et al. Exploring robust methods for evaluating treatment and comparison groups in chronic care management programs. Population Health Management. 2013;16:35–45. https://doi.org/10.1089/pop.2011.0104
9. International Federation of Red Cross and Red Crescent Societies. World Disasters Report 2016 Resilience: saving lives today, investing for tomorrow " " World Disasters Report Resilience: saving lives today, investing for tomorrow. 2016. Available from: www.ifrc.org.
10. WHO | Global status report on noncommunicable diseases 2014. WHO. 2015; Available from: http://www.who.int/nmh/publications/ncd-status-report-2014/en/.
11. Global Burden of Disease (GBD 2019) | Institute for Health Metrics and Evaluation. [cited 2021 Mar 26]. Available from: http://www.healthdata.org/gbd/2019.

12. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research. 2011;46:399–424. https://doi.org/10.1080/00273171.2011.568786
13. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. Statistics in Medicine. 2010;29:337–346. https://doi.org/10.1002/sim.3782
14. Setoguchi S, Schneeweiss S, Brookhart MA, et al. Evaluating uses of data mining techniques in propensity score estimation: A simulation study. Pharmacoepidemiology and Drug Safety. 2008;17:546–555. https://doi.org/10.1002/pds.1555
15. Olmos A, Govindasamy P. Propensity Scores: A Practical Introduction Using R | Journal of MultiDisciplinary Evaluation. Journal of MultiDisciplinary Evaluation. 2015;11. Available from: https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/431/414
16. Imbens GW. Matching Methods in Practice: Three Examples. IZA - Institute of Labor Economics. 2014;DP. https://doi.org/10.3386/w19959
17. Imbens G, Crump R, Hotz VJ, et al. Dealing with limited overlap in estimation of average treatment effects. Biometrika. 2009;96:187–199. https://doi.org/10.1093/biomet/asn055
18. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys. 2008;22:31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x
19. Cochran WG. The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. Biometrics. 1968;24:295. https://doi.org/10.2307/2528036
20. Iacus SM, King G, Porro G. Multivariate Matching Methods That Are Monotonic Imbalance Bounding. Journal of the American Statistical Association. 2011;106:345–361. https://doi.org/10.1198/jasa.2011.tm09599
21. Ripollone JE, Huybrechts KF, Rothman KJ, et al. Practice of Epidemiology Evaluating the Utility of Coarsened Exact Matching for Pharmacoepidemiology Using Real and Simulated Claims Data. This Am J Epidemiol. 2019;189:613–622. https://doi.org/10.1093/aje/kwz268
22. Su M, Zhou Z, Si Y, et al. Comparing the effects of China's three basic health insurance schemes on the equity of health-related quality of life: Using the method of coarsened exact matching. Health and Quality of Life Outcomes. 2018;16. https://doi.org/10.1186/s12955-018-0868-0
23. Schurer S, Alspach M, Macrae J, et al. The Medical Care Costs of Mood Disorders: A Coarsened Exact Matching Approach. 2015. https://doi.org/10.2139/ssrn.2564947
24. Pan W, Bai H. Propensity Score Analysis FUNDAMENTALS AND DEVELOPMENTS. Pan W, Bai H, editors. Guilford Press; 2015.
25. Rosenbaum PR, Rubin DB. The Bias Due to Incomplete Matching. Biometrics. 1985;41:103. https://doi.org/10.2307/2530647
26. Cochran WG, Rubin DB. Controlling Bias in Observational Studies: A Review on JSTOR. Journal of Statistics, Series A. 1973;35:417–446. Available from: http://www.jstor.org/stable/25049893
27. Health M of. 2008 STEPwise Approach to Chronic Disease Risk Factor Survey Report. Available from: https://www.who.int/ncds/surveillance/steps/2007_Report_StKitts.pdf
28. Geneva. Waist Circumference and Waist-Hip Ratio: Report of a WHO Expert Consultation [Internet]. Available from: www.who.int
29. Aryal K, Neupane S, Mehata S, et al. Non communicable diseases risk factors: STEPS Survey Nepal 2013 [Internet]. Katmandu: Kathmandu: Nepal Health Research Council; 2014. Available from: https://www.who.int/ncds/surveillance/steps/2012-13_Nepal_STEPS_Report.pdf
30. Morgan CJ. Reducing bias using propensity score matching. Journal of Nuclear Cardiology. 2018;25:404–406. https://doi.org/10.1007/s12350-017-1012-y
31. Gentleman R, Ihaka R. The R Project for Statistical Computing [Internet]. [cited 2017 Jun 12]. Available from: https://www.r-project.org/
32. Ho DE, Imai K, King G, et al. MatchIt: Nonparametric preprocessing for parametric causal inference. Journal of Statistical Software [Internet]. 2011;42:1–28. https://doi.org/10.18637/jss.v042.i08
33. King G, Nielsen R, King G, et al. Why Propensity Scores Should Not Be Used for Matching. Political Analysis. 2019;27:435–454. https://doi.org/10.1017/pan.2019.11