Working Paper Series, N. 5, May 2020

Department of Statistical Sciences
University of Padua
Italy

Evaluating inverse propensity score weighting in the presence of many treatments. An application to the estimation of the neighbourhood effect

Margherita Silan
Department of Statistical Sciences
University of Padua
Italy

Bruno Arpino
Department of Statistics, Computer Science, Applications
University of Florence
Italy

Giovanna Boccuzzo
Department of Statistical Sciences
University of Padua
Italy

Abstract: In this paper we consider the problem of estimating causal effects in a framework with many treatments through a simulation study. We engage in Monte Carlo simulations to evaluate the performance of inverse probability of treatment weighting (IPTW) with 10 treatments, estimating the propensity scores using Generalised Boosted Models. We assess the performance of IPTW under three different scenarios representing treatment allocations, and compare it with a simple parametric approach, i.e., logistic regression. IPTW's estimates are less biased, even though they exhibit a higher variance than those based on logistic regression. Moreover, we apply IPTW to the estimation of the neighbourhood effect on the probability of older people experiencing hospitalised fractures by comparing 10 neighbourhoods in the city of Turin (Italy). Our paper demonstrates that IPTW can be successfully applied to the estimation of neighbourhood effects, and, more generally, to the estimation of causal effects in presence of many treatments.

Keywords: Inverse probability of treatment weighting; Generalised Boosted Model; multi-treatment; neighbourhood effect

Final version (2019-08-13)

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Margherita Silan
silan@stat.unipd.it

Evaluating inverse propensity score weighting in the presence of many treatments. An application to the estimation of the neighbourhood effect

Margherita Silan
Department of Statistical Sciences
University of Padua
Italy

Bruno Arpino
Department of Statistics, Computer Science, Applications
University of Florence
Italy

Giovanna Boccuzzo
Department of Statistical Sciences
University of Padua
Italy

Abstract: In this paper we consider the problem of estimating causal effects in a framework with many treatments through a simulation study. We engage in Monte Carlo simulations to evaluate the performance of inverse probability of treatment weighting (IPTW) with 10 treatments, estimating the propensity scores using Generalised Boosted Models. We assess the performance of IPTW under three different scenarios representing treatment allocations, and compare it with a simple parametric approach, i.e., logistic regression. IPTW's estimates are less biased, even though they exhibit a higher variance than those based on logistic regression. Moreover, we apply IPTW to the estimation of the neighbourhood effect on the probability of older people experiencing hospitalised fractures by comparing 10 neighbourhoods in the city of Turin (Italy). Our paper demonstrates that IPTW can be successfully applied to the estimation of neighbourhood effects, and, more generally, to the estimation of causal effects in presence of many treatments.

Keywords: Inverse probability of treatment weighting; Generalised Boosted Model; multi-treatment; neighbourhood effect

1 Introduction

Propensity score techniques represent a way to simulate a randomized trial with observational data, when the use of a randomized controlled trial is not feasible and ethical. In a randomized trial the treatment allocation process is completely known and individuals' characteristics do not confound estimates. Differently from other

techniques in which the analyst models the outcome conditioning on all measured confounders, propensity score approaches are focused on modelling the treatment allocation process in order to make it 'known' as in a randomized trial. However, the treatment allocation model needs to be well specified, this is not trivial especially in the presence of many treatments.

The neighbourhood effect is the independent causal effect of living in a given neighbourhood rather than in another place on a given health or social outcome (Oakes, 2004). In its estimation a randomised experiment that randomly allocates individuals to different neighbourhoods should be performed. This approach would allow for the comparison of the health outcomes of individuals living in different neighbourhoods. Randomised experiments are, however, expensive and difficult to implement in the field of neighbourhood effects on health.

In observational neighbourhood studies, individuals self-select into different treatments (i.e., neighbours). For this reason, it is unclear whether differences in the outcomes of neighbourhoods can be causally attributed to living in one area instead of another, or whether these differences are simply due to the heterogeneous composition of the neighbourhoods - or in other words, to differences in the distribution of the characteristics of individuals living in different areas (Harding, 2003). Indeed, if the characteristics of these individuals not only vary across neighbourhoods, but are also associated with the outcome under study, they can be considered confounders of the neighbourhoods' effects.

To adjust for the observed characteristics of individuals, previous studies have often used parametric regression models, and, in particular, multilevel models with individuals nested into neighbourhoods. In the causal inference literature, it has been shown that regression models can be helpful in adjusting for observed confounders. But if groups differ greatly, these models may provide biased estimates due to extrapolation, which can be sensitive to model misspecification (Li et al., 2013; Drake, 1993).

As an alternative to parametric regression models, we propose dealing with the estimation of neighbourhood effect by using an inverse probability of treatment weighting (IPTW) approach that models the assignment to treatments (neighbourhoods) and the health outcome separately (Austin, 2011; Rosenbaum, 1987). In the first step, the method consists of estimating the probability of receiving each treatment (living in different neighbourhoods). In the second step, in estimating the outcome effect model, observations are weighted by the inverse of the probability of being treated. Recent studies have demonstrated the advantages of using non-parametric machine learning methods (Cannas and Arpino, 2019; Tu, 2019) to estimate the probability of receiving each treatment, defined as propensity scores. We follow these advances in the literature by using Generalised Boosted Models (GBM) (McCaffrey et al., 2013).

Previous methodological studies have shown that, compared to regression approaches, propensity score-based methods, and IPTW in particular, reduce the bias of causal estimates by guaranteeing a better balance of observed confounders across treatment groups (Austin, 2011; Austin and Stuart, 2015; Rosenbaum, 1987). Implementing IPTW for the estimation of neighbourhood effects is complicated by the fact that the number of neighbourhoods (treatments) to be compared is usually high.

IPTW has been employed and evaluated for a limited number of treatments (two or three). In this study, we evaluate and apply IPTW to many treatments.

Motivated by the estimation of neighbourhood effects on the probability of older people in the Italian city of Turin experiencing hospitalised fractures, we first run a series of Monte Carlo simulations to evaluate the performance of IPTW in the case of many treatments (10 specifically, which corresponds to the number of neighbourhoods in the real data). This simulation exercise is essential for gauging the feasibility of the approach in the context of neighbourhood observational studies, and, more generally, in the presence of many treatments. The simulations are also intended to evaluate the statistical performance in terms of the bias, variance, and coverage of IPTW with propensity scores estimated through GBM, and compared to a simpler parametric approach based on a logistic regression with neighbourhoods as the main independent variables.

The motivating case study refers to the estimation of the neighbourhood effect on the probability of older people experiencing hospitalised fractures, adjusting for confounders. The research question is to what extent observed differences in the incidence of hospitalised fractures across neighbourhoods can be causally attributed to the neighbourhoods' effects, rather than to their different compositions, i.e., to the fact that individuals with different risks factors for fractures live in different areas. Thus, we prefer to use a methodological approach that handles separately the assignment to treatments (neighbourhoods) and the occurrence of the health outcome, such as the IPTW (McCaffrey et al., 2013).

The rest of the paper is organised as follows. In the second section, we describe the IPTW approach and the GBM used to estimate the propensity scores. In the third section, we describe the motivating case study and the real data that inspired the simulations. In the fourth section, we describe the simulation study. We report the results of the study in section 5. In section 6, we illustrate the application on the real data. In section 7, we summarise the main findings and discuss possible future developments.

2 Methods

2.1 Generalized propensity score in a multi-treatment framework

Let suppose there is a population composed of N individuals, each of them indexed by $i = 1, \ldots, N$. Two fundamental variables are associated with each subject: a multivalued variable T that represents the treatment assignment and the outcome variable Y. It is possible to represent, for simplicity, the treatment assignment with a set of dummies $D_{it}(T_i)$ (Linden et al., 2016), where T_i is a multivalued treatment variable that takes values from 1 to K (in our specific application, it takes values from 1 to 10):

$$D_{it}(T_i) = \begin{cases} 1 & if \quad T_i = t \\ 0 & otherwise. \end{cases} \quad for \quad t = 1, \ldots, K \tag{1}$$

Consequently, we will have a set of potential outcomes $\mathbf{Y} = (Y_{1i}, ..., Y_{Ki})$ for individual i considering all different treatments, and just one of them is observed.

In order to apply propensity score techniques, some assumptions are needed, such as *temporality*, which implies that the treatment selection T must occur before the outcome; the *strong ignorability*, which is composed of two assumptions, unconfoundedness and positivity; and the *stable unit treatment value assumption (SUTVA)*. The strong ignorablity assumption requires that

- $Pr[\mathbf{Y}|T = t, x] = Pr[\mathbf{Y}|x]$, unconfoundedness assumption; and
- $0 < Pr[T = t|x] \quad \forall t \in T$, positivity assumption.

In other words, the potential outcomes, $\mathbf{Y}$, are independent of the treatment assignment T, given a set of observable variables X that are not affected by the treatment and each subject must have a positive probability to be included in all the treatment groups. The SUTVA includes two assumptions: the *no interference* and the *stable treatment assumption*. According to the SUTVA, the potential outcomes for any given unit do not vary with the treatments assigned to other units; and, for each unit, there are no different forms or versions of each treatment level that lead to different potential outcomes (Imbens and Rubin, 2015).

Imbens (2000) proposed a modification of the Rosenbaum-Rubin definition of the propensity score. The generalised propensity score (GPS) is the conditional probability of receiving a particular level of the treatment given the pre-treatment variables. In the literature, there are some scattered applications of GPS methods in multi-treatment frameworks, including a few applications in three (or four) treatments regimes (Tu and Koh, 2016). The most common model for estimating a GPS is the multinomial logistic regression (Lopez and Gutman, 2017), which produces K propensity scores e_{it} with $t = 1, ...K$, one for each treatment, that sum to 1. However, in this work, we adopted an approach proposed by McCaffrey et al. (2013) for a multi-treatment framework. This method is based on a GBM for computing the propensity score while reducing the risk of the misspecification of the treatment assignment model; and it is implemented in the `twang` package in R (Ridgeway et al. (2006), Toolkit for Weighting and Analysis of Nonequivalent Groups). The methodological research question consists in the evaluation of this approach in the presence of a high number of treatments, that has not yet been explored in the literature.

2.2 IPTW in a multi-treatment framework

The first step of the algorithm proposed by McCaffrey and colleagues (2013) consists of estimating the propensity score as in a dichotomous framework, while considering the treatment groups separately. For each treatment group t, the GBM fits a piecewise constant model composed of many simple regression trees in order to predict the dichotomous treatment (represented by the variable $D_t(T)$). These regression trees are combined to iteratively adjust the log-odds of treatment assignment $g(\mathbf{X})$ in order to maximise the log-likelihood function:

$$\ell(g) = \sum_{i=1}^{N} D_{it}(T_i)g(\mathbf{X}_i) - \sum_{i=1}^{N} log\{1 + exp[g(\mathbf{X}_i)]\} \tag{2}$$

where $D_{it}(T_i)$ is the treatment assignment indicator and $\mathbf{X}$ contains all the confounders (McCaffrey et al., 2004). The iterative process continues until the stopping rule is satisfied; in this case, it regards the balance of pre-treatment covariates. A possible balance measure is the Population Standardized Bias (PSB) for each variable v and each neighbourhood t. This measure compares the distribution of the confounders in each treatment group and in the whole population (p), while considering all treatment groups. It is given by the formula:

$$PSB_{vt} = \frac{|\hat{\bar{X}}_{vt} - \hat{\bar{X}}_{vp}|}{\hat{\sigma}_{vp}} * 100, \tag{3}$$

where $\bar{X}_{vt}$ is the mean of variable v computed on the analysed sample weighted with the inverse of the propensity score of being in neighbourhood t, and $\bar{X}_{vp}$ and $\hat{\sigma}_{vp}$ are the unweighted mean and the standard deviation of variable v in the whole population (McCaffrey et al., 2013).

The PSB balance measure is computed automatically for each variable and each neighbourhood, and needs to be summarised. In the `twang` package, it is possible to choose between two summary statistics: namely, the mean or the maximum value of PSB_{vt} among all considered covariates and treatment groups.

As we have a lot of dichotomous variables in our analysis, we have decided to use the Population Standardized Bias to measure the balance among the covariates; and, to be more conservative, to summarise it by its maximum value (instead of using the mean) among the pre-treatment variables. Indeed, minimising the maximum PSBs guarantees that all the other values are smaller than the maximum, whereas if we use the mean for the minimisation, there is the risk to have high values of the PSB offset by low values.

The R function `twang` allows us to set other important parameters such as the maximum number of trees to be combined (to reduce the risk of over-fitting), their maximum interaction level, and the shrinkage level. In this work, we used mainly default values of the function `mnps` in the `R` package `twang`; except for the following cases, in which we also followed other suggestions found in the literature (McCaffrey et al., 2004) (the results of these additional attempts are available from the authors):

- The number of GBM iterations (`n.trees`): we used the default value (10,000) for the empirical application, but, since we observed that the balance was reached with fewer iterations in the simulations, we set it at 3,000 in order to save time and computational effort. However, we also ran some simulations with 5,000, 10,000, and 20,000 GBM iterations in order to check whether the lower number negatively affected the performance of IPTW.

- A shrinkage parameter was applied to each tree in the expansion (`shrinkage`). The default value was 0.01, but we also used 0.0005 in some simulations, as suggested in the literature (McCaffrey et al., 2004).

- For the fraction of the training set, observations were randomly selected to propose the next tree in the expansion (`bag.fraction`), while introducing randomness into the model fit if it was less than 1. The default value was

1, but we also used 0.5 in some simulations, as suggested by the literature (McCaffrey et al., 2004).

- For the maximum number of iterations for the direct optimisation (`iterlim`): the default value was 1,000, but we also tried some simulations with a higher value (10,000) to check whether 1,000 was enough.

After the GBM computation of the propensity scores for each individual and for each treatment with respect to the rest of the population has been implemented, the result is a matrix with K propensity scores for each individual, each one of which is referred to as one of the treatments. In other words, the first part of the algorithm produces a matrix that shows the computed probability of living in each neighbourhood (and not in other neighbourhoods) for each subject. This step produces propensity scores that are useful for making each treatment group comparable with the rest of the population. The sum of the K propensity scores for each individual is not equal to 1, as in the multinomial model, because these values are the results of different models that consider treatments separately. Since propensity scores are used in this work primarily in order to balance the weights, this is not an issue, and it is not necessary to modify these values to make them sum to 1. Indeed, such a transformation would simply modify the scale of the weights, and would not have any effect on the final result. The final weight w_i for each subject is given by the inverse of the propensity score e_{it} of the received treatment

$$w_i = \sum_{t=1}^{K} \frac{D_{it}(T_i)}{e_{it}}. \tag{4}$$

Once weights are computed, several estimands may be considered for the estimation of the treatment effect. In this work we considered the Average Treatment Effect (ATE) that is defined as

$$\hat{ATE}_{t',t''} = \frac{1}{N}\sum_{i=1}^{N} \frac{Y_i D_{it'}(T_i)}{e_{it'}} - \frac{1}{N}\sum_{i=1}^{N} \frac{Y_i D_{it''}(T_i)}{e_{it''}}. \tag{5}$$

considering two treatments t' and t''.

When dealing with IPTW, it is common to find extremely high weights that cause the variance of estimates to increase. Therefore, weight trimming has been considered as a way to reduce the variance with small losses in terms of bias (Lee et al., 2011). However, the optimal level of trimming for improving the inference and achieving the best compromise between bias and variance is difficult to determine. Thus, it is sometimes more effective to focus on the procedure for computing weights, such as a proper specification of the propensity score model (Lee et al., 2011). Nonetheless, we also implemented an asymmetrical trimming of the higher weights in the simulation study, while setting the extreme weights equal to the upper bound threshold, even if there is no proof of a substantial improvement in the overall performance of the GBM in an IPTW procedure in dichotomous cases (Lee et al., 2011).

3 Motivating case study

As our motivating case study, we consider the estimation of the neighbourhood effect on the incidence of hospitalised fractures among older people living in the Italian city of Turin. Previous studies have found that the neighbourhood context may affect fracture rates in the residents through two main paths: the terrain may be uncomfortable to walk on, which can cause people to fall; or the people living in the area may be discouraged from engaging in physical activity, which can lead to a deterioration in their muscle and bone mass (Sànchez-Riera et al., 2010; Barnett et al., 2017) .

Data used in the analysis come from the Longitudinal Study in Turin. This is an integrated database that includes administrative data flows with information about the residents from both censuses and health data flows (hospital discharge records, participation in prescription charges, and territorial drug prescriptions). These data sources may be linked together and through time (starting from 1971) using a deterministic key that is unique for each individual who was a resident of Turin for at least one day.

The analysed population consists of all participants in the 2001 population census, with some additional restrictions. We consider only the individuals who where aged 60 or older on 31 December 2001. In order to be able to collect information on possible confounders related to past health information, we focus on the individuals who were living in Turin between 1 January 1997 and 31 December 2001. Finally, we measure the outcome, i.e., the incidence of hospitalised fractures during the year following the census (2002). Therefore, we restrict our analyses to individuals who were living in Turin over the whole period between 1 January 1997 and 31 December 2002. Our design allows us to measure the time-varying confounders before the treatment, which is in turn measured before the outcome is observed. The final population counts 225,828 individuals that are not equally distributed among different neighbourhoods.

In the application, we focus on the causal effect of living in a given neighbourhood at the 2001 census on the probability of experiencing at least one hospitalised fracture during 2002.

The city of Turin can be divided into 10 neighbourhoods, which have different living conditions (e.g., levels of deprivation, walkability, crime, and social cohesion) and population characteristics.

Based on the literature about neighbourhood effects on older people's health (Roux et al., 2004; Yen et al., 2009), we consider the following variables as possible confounders: gender, age, region of birth, family composition, educational attainment, last observed professional condition, home ownership, and overcrowding. The region of birth has been coded by distinguishing between those born in Piedmont (the region to which Turin belongs); in another region in the north of Italy; in the centre of Italy; in the south of Italy or the islands; or in a country outside of Italy. The variable that represents the family composition is built by combining the individual's marital status and family components: living alone; married and living with the partner only (two components); not married and not living alone (two or more components); and married and living in a family with more than two peo-

ple. The variable that reflects the individual's last observed professional condition is composed based on census data from 1971 to 2001, with the aim of capturing the person's last type of employment before retirement. For some individuals this was not possible because they were already retired in 1971 (or in all of the censuses they were observed) or they were not working for other reasons. Additionally, the professional status variable distinguishes between this group and homemakers, entrepreneurs, white-collar workers, and manual workers. The variable representing overcrowding consists of the ratio between the number of rooms and the number of family components. Moreover, in the simulation study two more variables that describe the health conditions of the individuals in the neighbourhoods are reported: diagnoses of hypertension or cardiac issues and the number of different kinds of drugs that have been prescribed to individuals. We do not use these variables in the empirical study because they can themselves be affected by the treatment. However, they are included in the set of variables that are considered in the simulation study because in that context we are manipulating the true data-generating models.

Some descriptive statistics on the outcome and the confounders considered in our empirical analyses by neighbourhood may be found in appendix A, table 5.

4 Simulation design

In order to keep our experiment realistic and to simplify our computations, we extracted from the total population a 10% sample from each neighbourhood. The original data structure was thereby preserved, but with a reduced sample size that makes the computations less demanding (the simulation dataset contains 22,690 individuals). In order to keep the simulation simple, we selected a small number of covariates from the variables described in section 3: gender, age, education (Edu0, Edu1, Edu2, and Edu3), overcrowding, hypertension, and drugs.

In the simulation experiment, we included variables describing the health conditions of the population that we had discarded in the empirical study, because in our simulations we established both the temporality and the causality direction given by the data generation design: i.e., we simulated first the treatment and then the outcome; whereas in the empirical framework, this assumption could not be fully trusted with respect to health conditions.

In line with other studies (Arpino and Cannas, 2016; Setoguchi et al., 2008), and given the real distribution of these six covariates, we decided to simulate the treatment assignment and the outcome according to three different scenarios that reflect three different treatment allocation settings: the first one reflects the real circumstances with a simple, linear, and additive model; the second one shows a case in which the treatment allocation equation is complex and may be misspecified; and the third one represents a highly unbalanced situation.

Simulations have been implemented with the software `R`, the code for the simulation is reported in the appendix D.

In the first scenario, the treatment assignment equation is simple and close to reality. The treatment is generated through a multinomial logistic model, using neighbourhood 6 as a reference; because it is, the neighbourhood with the lowest

crude hospitalised fractures rate. Thus, for each neighbourhood t and each individual i, the treatment equation is

$$\begin{aligned} ln\left(\frac{Pr(T_i=t)}{Pr(T_i=6)}\right) &= {}_1\beta_0^t +_1 \beta_1^t * Gender_i +_1 \beta_2^t * Age_i +_1 \beta_3^t * Edu1_i + \\ &+ {}_1\beta_4^t * Edu2_i +_1 \beta_5^t * Edu3_i +_1 \beta_6^t * Overcrowding_i \\ &+ {}_1\beta_7^t * Hypertension_i +_1 \beta_8^t * Drugs_i. \end{aligned} \quad (6)$$

In order to choose the values for the coefficients, we estimated a multinomial logistic model on the whole population and used the same rounded parameters for $t = 1, ..., 5, 7, ..., 10$, for the intercept, ${}_1\beta_0^t$, and for other coefficients, ${}_1\beta_v$ $v = 1, ...8$ (the exact values of the parameters are reported in table 6 in appendix B).

The second scenario relies on a more complex treatment assignment equation that includes six interaction terms and three quadratic terms, while having the following equation form for each neighbourhood t

$$\begin{aligned} ln\left(\frac{Pr(T_i=t)}{Pr(T_i=6)}\right) &= {}_2\beta_0^t +_2 \beta_1^t * Gender_i +_2 \beta_2^t * Age_i +_2 \beta_3^t * Edu1_i \\ &+ +_2\beta_4^t * Edu2_i +_2 \beta_5^t * Edu3_i +_2 \beta_6^t * Overcrowding_i + \\ &+ {}_2\beta_7^t * Hypertension_i +_2 \beta_8^t * Drugs_i +_2 \beta_9^t * Age_i^2 + \\ &+ {}_2\beta_{10}^t * Overcrowding_i^2 +_2 \beta_{11}^t * Drugs_i^2 + \\ &+ {}_2\beta_{12}^t * Gender_i * Age_i +_2 \beta_{13}^t * Gender_i * Hypertension_i + \\ &+ {}_2\beta_{14}^t * Gender_i * Drugs_i +_2 \beta_{15}^t * Age_i * Hypertension_i + \\ &+ {}_2\beta_{16}^t * Age_i * Drugs_i +_2 \beta_{17}^t * Drugs_i * Hypertension_i. \end{aligned} \quad (7)$$

As in the first scenario, the parameters for these treatment assignment equations were chosen based on the parameters estimated by a multinomial logistic model with the same functional form for the whole population (the exact values of the parameters are reported in table 7 in appendix B).

The third scenario relies on the very same treatment assignment equation as in the first scenario, but with different parameters. Indeed, starting with the coefficients in scenario 1, some of the parameters were modified to obtain a greater initial imbalance. Moreover, in order to keep the simulated dataset close to a potentially real situation in terms of the hospitalised fractures percentage, the intercepts were modified as well (the exact values of the parameters are reported in table 8 in appendix B).

We evaluated the initial balance of these three scenarios in all of the 1000 simulations using the Population Standardized Bias. The mean values of the PSB among all of the 1000 simulations for each scenario are reported in table 1. While in the first scenario the initial situation is only mildly unbalanced, in scenarios 2 and 3 more extreme imbalanced situations can be observed.

After the treatment generation, the outcome has also been simulated given the

Table 1: Mean of PSB among the neighbourhoods of the unweighted sample in all of the iterations (simulation study).

		Neighbourhoods									
Scenario	**Variable**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	Male	4.64	1.33	1.59	2.58	2.33	2.49	1.68	2.33	1.61	5.02
	Female	4.64	1.33	1.59	2.58	2.33	2.49	1.68	2.33	1.61	5.02
	Age	17.84	5.61	5.11	6.88	8.46	9.47	3.80	12.70	2.61	16.65
	Primary Educ. or lower	36.04	10.56	6.76	6.53	22.68	23.36	2.24	26.42	1.35	26.14
	Lower Secondary Educ.	8.40	5.94	3.04	1.74	2.68	3.29	1.30	3.76	5.42	4.94
1	Upper Secondary Educ.	19.90	6.99	4.81	5.79	16.12	17.14	1.86	15.57	2.40	18.54
	Tertiary Educ.	31.30	1.56	0.83	1.14	15.87	14.45	1.27	20.77	5.61	15.36
	No Hypertension	7.41	1.48	1.83	1.77	2.22	5.29	1.91	4.23	1.80	3.95
	Hypertension	7.41	1.48	1.83	1.77	2.22	5.29	1.91	4.23	1.80	3.95
	Overcrowding	26.34	6.00	2.19	1.36	12.02	10.18	2.54	21.93	4.44	2.32
	Drugs	24.53	1.71	5.29	4.04	10.31	12.92	2.14	13.22	3.99	14.71
	Male	6.26	1.41	1.64	6.05	1.42	6.38	6.22	3.76	2.78	1.59
	Female	6.26	1.41	1.64	6.05	1.42	6.38	6.22	3.76	2.78	1.59
	Age	30.81	6.14	13.49	29.99	2.49	14.87	26.13	26.09	38.62	1.58
	Primary Educ. or lower	29.91	9.77	10.60	8.64	23.39	14.35	2.73	21.47	4.02	20.31
	Lower Secondary Educ.	3.42	5.39	4.00	2.78	3.14	1.89	2.67	3.22	5.84	7.39
2	Upper Secondary Educ.	15.61	6.47	6.36	7.04	15.14	10.56	2.21	13.87	5.84	12.81
	Tertiary Educ.	24.22	1.21	2.25	2.45	18.31	14.96	2.33	15.25	7.51	6.26
	No Hypertension	1.39	3.94	5.45	6.98	1.79	3.30	5.22	9.65	7.74	2.51
	Hypertension	1.39	3.94	5.45	6.98	1.79	3.30	5.22	9.65	7.74	2.51
	Overcrowding	19.63	9.68	2.65	10.62	17.49	11.54	11.88	14.70	15.59	1.00
	Drugs	13.52	11.52	9.99	11.93	3.85	2.52	15.86	17.31	7.14	4.79
	Male	22.94	3.75	6.95	38.93	8.38	5.34	1.55	13.88	7.79	5.61
	Female	22.94	3.75	6.95	38.93	8.38	5.34	1.55	13.88	7.79	5.61
	Age	258.46	23.91	17.88	7.20	27.36	23.28	15.68	20.03	18.78	13.86
	Primary Educ. or lower	4.19	22.26	23.94	9.30	27.75	26.14	10.24	52.42	3.44	31.52
	Lower Secondary Educ.	6.70	33.45	10.18	4.36	2.88	2.24	6.12	27.87	8.68	5.90
3	Upper Secondary Educ.	1.12	4.59	44.43	1.15	21.08	20.27	15.01	10.57	5.85	22.65
	Tertiary Educ.	2.29	12.91	13.03	9.53	21.48	19.93	8.85	92.91	13.00	21.24
	No Hypertension	28.29	3.94	3.70	2.42	1.16	3.22	7.56	10.45	1.59	13.86
	Hypertension	28.29	3.94	3.70	2.42	1.16	3.22	7.56	10.45	1.59	13.86
	Overcrowding	34.09	6.35	5.08	3.67	30.76	8.17	3.47	21.79	4.21	4.67
	Drugs	27.08	4.81	11.89	2.72	5.21	9.27	8.47	28.43	2.24	44.81

six covariates and the treatment assignment according to the following model:

$$\begin{aligned} ln\left(\frac{Pr(Y_i=1)}{Pr(Y_i=0)}\right) &= \beta_0 + \beta_1 * Gender_i + \beta_2 * Edu1_i + \beta_3 * Edu2_i + \\ &+ \beta_4 * Edu3_i + \beta_5 * Hypertension_i + \beta_6 * Age_i + \\ &+ \beta_7 * Overcrowding_i + \beta_8 * Drugs_i + \beta_9 * D_{i1}(T_i) + \\ &+ \beta_{10} * D_{i2}(T_i) + \beta_{11} * D_{i3}(T_i) + \beta_{12} * D_{i4}(T_i) + \\ &+ \beta_{13} * D_{i5}(T_i) + \beta_{14} * D_{i7}(T_i) + \beta_{15} * D_{i8}(T_i) + \\ &+ \beta_{16} * D_{i9}(T_i) + \beta_{17} * D_{i10}(T_i), \end{aligned} \tag{8}$$

where $D_{i1}(T_i)$, $D_{i2}(T_i)$, ..., $D_{i10}(T_i)$ are dichotomous variables that take value 1 if the individual i lives in the considered neighbourhood, and value 0 otherwise. As before, the reference is neighbourhood 6. The coefficients are close to those estimated by the same model for the whole population, but the parameters from β_9 to β_{17} were

inflated to obtain a larger neighbourhood effect for the purposes of estimation (the exact values of the parameters are reported in table 9 in appendix C). Indeed, when the true neighbourhood effects are small, there is a risk that the simulations will produce more biased and less stable estimates, and that the IPTW approach will perform badly (Cepeda et al., 2003). However, we also ran some simulations with smaller neighbourhood effects in order to explore and verify this result in a multi-treatment framework.

We evaluated the performance of the two approaches, the logistic regression, and the IPTW, while comparing the estimates of nine neighbourhood coefficients (the reference is neighbourhood 6) and the true treatment effect used to simulate the outcome. The analysis was focused on three measures: the mean and the median of the relative bias (the percentage difference from the true treatment effect), the variance of the estimated values among the 1000 simulations, and 95% confidence interval coverage (the percentage of times the true value is included in the 95% confidence interval of the obtained estimates among all of the simulations).

5 Simulation results

For each replicate in every scenario, we estimated the neighbourhood effect using both the logistic regression approach and the IPTW approach. Since we were trying to improve the balance of the confounders among the neighbourhoods, we observed the distribution of the PSB across all of the simulations, neighbourhoods, and variables. To summarise them all, we reported the mean of the PSB of the weighted samples among all of the simulations in table 2.

In the first two scenarios, the balance attained with IPTW was extremely good, with all of the considered confounders showing an average PSB that was lower than 5% for all of the neighbourhoods; indeed, in many cases, the PSBs was even lower than 1%. According to the literature, the possible thresholds for defining a balanced population are 25%, 20%, and 10% (Austin, 2009; Rosenbaum and Rubin, 1985). Using even the most restrictive threshold cited in literature, we can state that in these two scenarios the balance was reached.

In the most complicated scenario (scenario 3), the PSBs tended to be higher. This was especially the case for neighbourhood 1, for which most of covariates had PSBs higher than 10%, and the average PSB for age was 53.23%. Even though the balance was not satisfactory, it should be noted that in scenario 3 the initial imbalance was very high (e.g., the PSB for age in neighbourhood 1 was 258.46; table 1). Indeed, if we compare the balance after weighting (table 2) with the initial balance, we can see that even in scenario 3, the use of the IPTW approach guarantees a considerable improvement in the degree of similarity of the confounders' distributions across the neighbourhoods. Since the residual imbalance was higher, we expected to observe higher bias for the IPTW estimator in scenario 3.

Whereas in scenarios 1 and 2 the bias of the IPTW estimates was quite good, or lower than 5% in most cases, in the third scenario there were two parameters with a bias higher than 10%. However, as was already explained, in the third scenario, the initial balance was particularly challenging in terms of the distribution

Table 2: Mean of PSB among the neighbourhoods of the weighted sample in all of the replicates (simulation study).

		Neighbourhoods									
Scenario	**Variable**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	Male	1.90	0.42	0.51	0.73	0.82	0.88	0.54	1.42	0.48	1.62
	Female	1.90	0.42	0.51	0.73	0.82	0.88	0.54	1.42	0.48	1.62
	Age	0.96	1.45	0.79	1.02	2.66	2.91	1.30	1.47	2.00	5.48
	Primary Educ. or lower	1.55	0.29	0.27	0.40	1.60	1.67	0.56	1.01	0.57	3.23
	Lower Secondary Educ.	0.83	0.40	0.32	0.40	0.53	0.51	0.38	0.84	0.72	0.80
	Upper Secondary Educ.	0.73	0.24	0.20	0.26	0.72	0.87	0.50	0.41	0.42	1.75
	Tertiary Educ.	0.35	0.53	0.27	0.33	2.38	2.26	0.55	0.27	1.45	4.04
	No Hypertension	0.93	0.42	0.31	0.44	0.65	1.01	0.74	0.91	0.64	1.53
	Hypertension	0.93	0.42	0.31	0.44	0.65	1.01	0.74	0.91	0.64	1.53
	Overcrowding	3.35	1.05	0.26	0.41	0.47	0.41	0.43	3.12	0.45	2.10
	Drugs	1.08	0.96	0.52	0.80	2.18	2.41	0.98	1.24	1.69	4.48
2	Male	1.04	0.42	0.85	2.67	0.67	0.70	1.20	2.35	1.73	0.60
	Female	1.04	0.42	0.85	2.67	0.67	0.70	1.20	2.35	1.73	0.60
	Age	0.52	0.53	1.87	6.13	1.73	2.88	4.59	6.21	7.87	2.01
	Primary Educ. or lower	1.07	0.35	0.62	1.23	1.64	1.62	1.24	1.56	2.35	1.22
	Lower Secondary Educ.	0.66	0.31	0.51	1.13	0.63	0.63	1.09	1.26	1.80	0.19
	Upper Secondary Educ.	0.38	0.22	0.38	0.91	0.59	0.71	1.29	0.86	2.00	1.02
	Tertiary Educ.	0.22	0.21	0.31	1.09	2.78	2.58	1.00	0.68	3.62	0.63
	No Hypertension	0.57	0.35	0.59	1.42	0.43	0.96	1.27	1.56	1.52	0.50
	Hypertension	0.57	0.35	0.59	1.42	0.43	0.96	1.27	1.56	1.52	0.50
	Overcrowding	1.33	1.93	0.50	0.94	0.48	0.40	0.79	3.22	1.23	0.48
	Drugs	0.57	1.18	0.63	2.23	1.64	1.97	1.64	2.28	3.55	0.97
3	Male	9.59	0.85	0.57	1.66	1.53	1.46	0.98	1.47	1.21	2.25
	Female	9.59	0.85	0.57	1.66	1.53	1.46	0.98	1.47	1.21	2.25
	Age	53.23	5.51	4.00	4.59	7.35	7.11	5.47	4.82	5.87	8.70
	Primary Educ. or lower	10.46	0.54	0.33	1.28	2.80	2.59	1.73	1.73	1.36	5.18
	Lower Secondary Educ.	8.80	1.30	0.58	0.61	0.73	0.81	0.69	0.88	1.13	1.13
	Upper Secondary Educ.	9.16	0.45	0.54	0.74	1.20	1.15	1.79	0.81	0.85	2.54
	Tertiary Educ.	7.07	2.19	1.80	2.32	4.02	3.86	1.45	0.70	2.76	6.74
	No Hypertension	11.19	0.61	0.54	0.78	0.78	0.97	0.99	1.54	0.73	2.61
	Hypertension	11.19	0.61	0.54	0.78	0.78	0.97	0.99	1.54	0.73	2.61
	Overcrowding	18.68	1.36	0.49	0.85	1.98	0.66	0.64	2.02	0.59	2.20
	Drugs	17.18	1.22	0.74	1.07	2.31	2.62	1.24	1.91	1.87	9.59

of the confounders among the different treatment groups. Moreover, when the bias of IPTW estimates was high, the logistic regression method also provided biased estimates.

In the first scenario, we observe that the biases relative to the estimates produced by IPTW were smaller than those produced by the logistic regression method, except for one neighbourhood, number 10. This neighbourhood had the highest PSBs, and was the only one for which a PSB higher than 5% was found. Relative to the other neighbourhoods, the third had the largest bias with respect to both the estimation approaches and the mean and the median. Indeed, the bias of this parameter was expected to be the highest because its true value was the smallest and closest to 0. According to the literature (Cepeda et al., 2003), higher bias is often observed for estimates of smaller effects. In general, in the first scenario, almost all of the

parameters were estimated by both of the models with bias lower than 5%; except for neighbourhoods 1, 3 (already mentioned), and 4, where, on average, the IPTW approach seems to have provided better estimates.

In the second scenario, the estimates given by the IPTW method for neighbourhoods 4, 8, and 9 had a particularly high median bias, of between 5% and 10%. This was probably because the balance in these neighbourhoods was not completely achieved, especially for the variable age, which had a mean PSB of more than 5% in these three neighbourhoods (table 2). On the other hand, the logistic regression model provided estimates that were particularly biased for the effect of neighbourhood 1; probably because of the initial highly unbalanced situation (as shown in table 1).

In the third scenario, both methods performed well for most of the neighbourhoods (numbers 2, 5, 7, 8, 9, and 10), as they had both mean and median biases of less than 5%. However, when the IPTW approach was applied, the biases became slightly smaller for all of these neighbourhoods. The logistic regression model produced better results in terms of bias in neighbourhoods 3 and 4. However, both methods performed poorly with respect to the estimates for neighbourhood 1, for which the initial situation was extremely unbalanced (as shown in table 1). While the situation of the first neighbourhood remained unbalanced even after weighting (as shown in table 2), the IPTW approach produced estimates for this parameter that had, on average, half the bias of those produced by the logistic regression model.

A general observation with respect to table 3 is about variances. Indeed, in all of the scenarios and for all of the neighbourhoods, the variances of the estimates generated by the IPTW approach were higher than those produced by the logistic regression model. In the first scenario, which corresponds most closely to reality, the variances of the two models were more similar and smaller than those in the other two scenarios, in which the allocation of individuals to treatments was more complex (in the second one) and more unbalanced (in the third one).

Since in the presence of weights the variance may increase and the estimates may be greatly affected, especially if the weights are extreme, we also tried an asymmetrical trimming. We trimmed only the extremely high weights, reducing the influence of those individuals who were under-represented in some of the neighbourhoods, based on the assumption that these individuals were outliers who did not reflect the population as a whole. Selecting different levels of trimming (percentiles from 99 to 85 and 75), we assigned the threshold weight value to those individuals who had higher weights. This technique proved to be quite useful for reducing the variance, but the gain was associated with an increase in bias in some cases.

In figures 1 and 2, two examples are presented that show how trimming affected the mean bias and the variance of the estimates produced by the IPTW approach relative to the mean bias and the variance of the logistic regression's estimates for neighbourhoods 5 and 8 in the first scenario. In some cases, as for neighbourhood 5 in figure 1, it was possible to increase the level of trimming while having a limited impact on the bias, or even causing it to decrease slightly at around the 85^{th} percentile, while ensuring that it remained lower than the bias of the estimates produced by the logistic regression. On the other hand, the variance of the trimmed estimates was substantially reduced, and assumed values closer to the variance of

Table 3: Variance, 95% Confidence Interval Coverage and bias (mean and median) of parameters' estimates in the 3 scenarios, comparison between regression and IPTW (simulation study).

		Scenario 1				**Scenario 2**				**Scenario 3**			
Neigh.		**Bias**		**Var**	**95% CI**	**Bias**		**Var**	**95% CI**	**Bias**		**Var**	**95% CI**
(*true value*)	**Method**	**Mean**	**Median**		**Coverage**	**Mean**	**Median**		**Coverage**	**Mean**	**Median**		**Coverage**
1	Logit	6.09	5.26	0.08	96.0	13.32	10.64	0.08	94.2	62.38	60.64	0.09	63.2
(0.820)	IPTW	3.90	3.74	0.10	96.1	6.28	3.93	0.09	95.4	37.16	27.75	0.29	88.3
2	Logit	2.05	1.29	0.07	95.3	2.85	1.20	0.07	96.2	3.09	2.13	0.10	95.5
(1.310)	IPTW	0.97	-0.15	0.08	95.0	1.67	0.33	0.07	96.4	2.66	0.77	0.15	93.0
3	Logit	11.92	11.96	0.08	95.8	3.28	3.67	0.11	95.0	12.97	9.80	0.11	95.9
(0.375)	IPTW	11.19	8.24	0.09	94.9	4.96	3.84	0.12	94.9	18.30	14.17	0.17	93.2
4	Logit	5.79	3.07	0.08	96.2	-4.87	0.32	0.22	96.6	3.75	2.76	0.11	95.2
(0.720)	IPTW	4.40	2.39	0.10	95.4	-9.53	-5.37	0.25	96.3	8.22	5.57	0.16	92.7
5	Logit	2.67	2.36	0.08	95.8	3.86	2.71	0.08	95.7	2.03	1.38	0.11	94.5
(0.915)	IPTW	1.59	1.41	0.09	95.0	1.43	0.08	0.09	94.1	1.16	-0.78	0.17	92.1
7	Logit	2.37	1.27	0.07	95.7	-0.71	-0.34	0.13	95.3	2.59	1.59	0.10	95.9
(1.430)	IPTW	0.71	-0.02	0.08	95.0	-4.25	-4.01	0.14	95.0	1.98	-0.02	0.15	92.5
8	Logit	3.66	4.33	0.09	96.5	-3.12	-0.28	0.21	94.5	5.18	3.81	0.12	95.6
(0.950)	IPTW	-0.12	-1.21	0.11	95.8	-7.18	-4.63	0.24	93.7	-0.29	-0.15	0.22	94.2
9	Logit	3.01	3.14	0.08	96.1	-5.78	-2.43	0.22	96.2	1.51	-0.58	0.11	96.4
(1.020)	IPTW	1.36	-0.32	0.09	95.7	-11.52	-8.50	0.26	95.1	1.82	-0.51	0.15	94.0
10	Logit	-0.54	-1.13	0.09	95.8	1.36	0.06	0.07	94.8	1.17	-0.51	0.11	94.8
(1.535)	IPTW	-3.68	-3.96	0.10	95.1	-0.59	-1.20	0.08	94.1	-0.53	-1.81	0.18	93.0

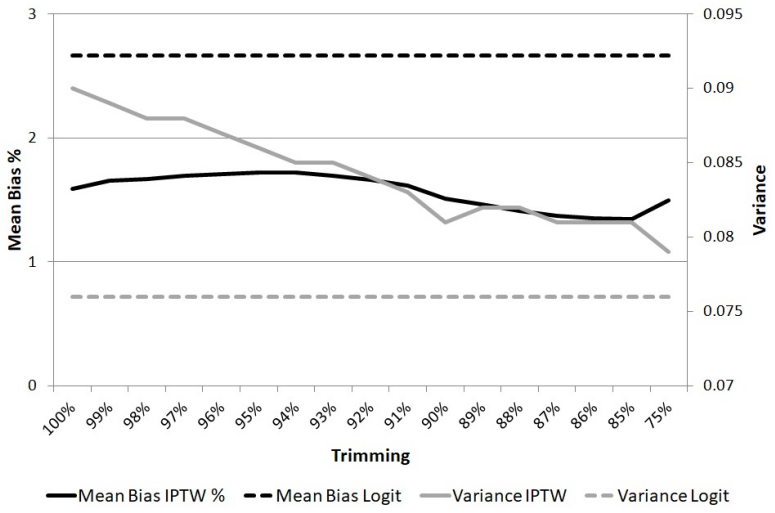

Figure 1: Comparison of the mean biases and the variances of the estimates obtained by IPTW at different levels of trimming for the neighbourhood effect of neighbourhood 5 in the first scenario with the logistic regression's (Logit) results (represented as horizontal dashed lines).

the logistic regression's estimates when the level of trimming was increased. Thus, when we consider this example, we can state that the optimal level of trimming in order to reduce both the bias and the variance may be around the 85^{th} percentile.

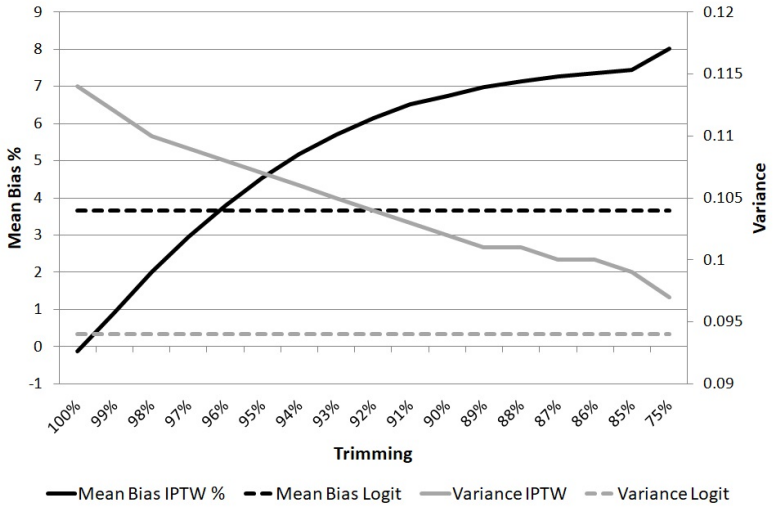

Figure 2: Comparison of the mean biases and the variances of the estimates obtained by IPTW at different levels of trimming for the neighbourhood effect of neighbourhood 8 in the first scenario with the logistic regression's (Logit) results (represented as horizontal dashed lines).

A completely different situation can be observed for the estimates of neighbour-

hood 8 in the first scenario, where the trade-off between the bias and the variance was more severe than in the previous example. Indeed, when the level of trimming was increased, the bias grew from around 0% in the absence of trimming to around 8% at the 75^{th} percentile of trimming. However, the variance decreased when moving closer to the variance of the logistic regression's estimates. Indeed, finding the optimal level of trimming was harder in this case, as the level at which the estimates were less biased was the one at which the variances were higher. Moreover, at around the 96^{th} percentile of trimming, we got estimates with the same bias as the logistic regression's estimates, but with a variance that was 15% higher.

In general, even after observing all of the trimmed estimates in all of the simulations, it was not possible to find a common criterion we could use to define a best practice in terms of trimming. The fact that we had nine different parameters to estimate did not make this choice easier, because the levels that ensure a balance between bias and variance may be different for each parameter. Moreover, in an empirical framework, it is not possible to observe the bias of estimates. Thus, it would have been even more difficult to discern which trimming level was the best to use without quantifying the loss in terms of bias. Thus, we would not recommend the trimming of weights when using the IPTW approach in a multi-treatment framework.

In order to improve the performance of the IPTW approach, we also tried to change some default settings in the `twang` package. As we mentioned before, we ran some simulations with different numbers of GBM iterations (3,000, 5,000, 10,000, and 20,000), levels of shrinkage (0.01 and 0.0005), fractions of the training set to fit the trees (1 and 0.5), and maximum numbers of iterations for the direct optimisation (1,000 and 10,000); as well as several combinations thereof. However, as the balance after weighting was not improved and the bias was not reduced, we decided against deepening this research path, and instead opted to use all of the default values for the simulations, except for the number of GBM iterations (the default was 10,000, but to save time and computational effort, we used 3,000, since the balance was reached with fewer iterations).

6 Empirical results

In this section, we describe the empirical results obtained by both the logistic regression and the IPTW approaches for the estimation of the neighbourhood effects of 10 neighbourhoods of Turin city. As we explained in section 3, we are considering the population aged 60 or older, with the hospitalised fractures event as the outcome. The confounders we consider are age, gender, region of birth, family composition, education level, last observed professional condition, home ownership, and overcrowding; which are described in detail in section 3.

As we explained in section 2, we estimated the weights for the IPTW approach using the default values of the function `mnps` in the `R` package `twang`, and included the `n.trees` number of GBM iterations that was set to 10,000, even if such a large number was not necessary. Indeed, as figure 3 shows, gains in balance become smaller after 3,000 iterations; and in some cases, such as in neighbourhood 7, increasing the

complexity of the GBM model may worsen the obtained balance of the weighted variables and cause overfitting. However, when a huge number of treatments is to be considered, the decision about what number of iterations is optimal for getting good results in correspondence with each of them it is not trivial.

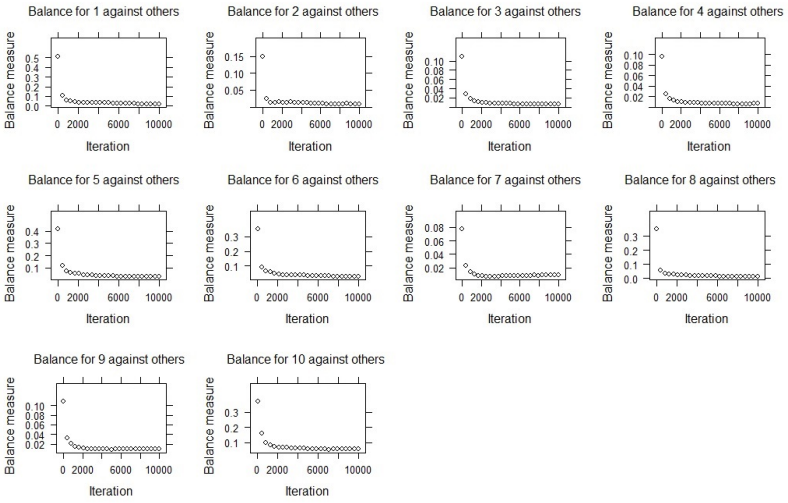

Figure 3: Reduction of the balance measure (the Population Standardized Balance, computed as in equation 3) during the weights estimation process in correspondence with an increasing number of iterations for the 10 considered neighbourhoods, and considering the whole population.

Nevertheless, as shown in figure 4, the final result is quite satisfying with respect to the initial unweighted situation, with almost all of the significant reductions of the PSB differences considering the maximum among all pairwise comparisons.

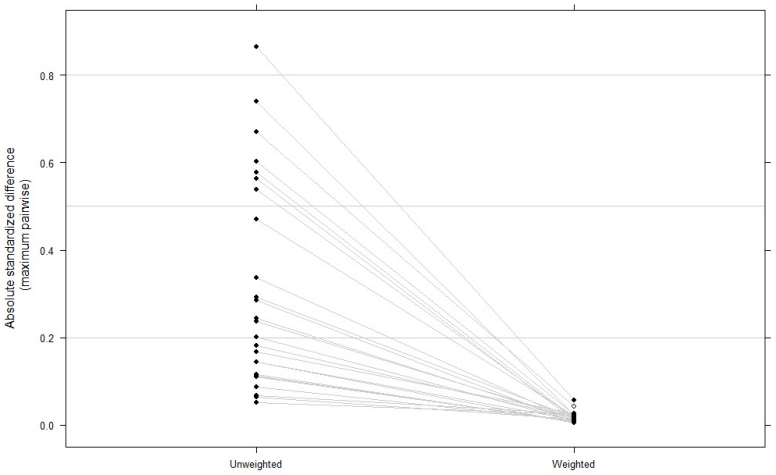

Figure 4: Comparisons of the absolute standardised differences (considering the maximum of pairwise comparisons) in the whole population before and after weighting.

There are small differences between the logistic regression and the IPTW estimates. We computed the neighbourhood effect for the whole population (All), and for the female (Women) and male (Men) populations separately. In table 4, we report the neighbourhood effect estimates in terms of the odds ratio, with standard errors and 95% confidence intervals for the two estimation approaches, and with respect to the three different populations selected.

The parameters estimated with the two approaches on the whole population were similar, except for neighbourhood 10, for which the effect was greater in the logistic regression model. The odds for the individuals living in neighbourhoods 2 and 7 was 33% higher than for the people living in neighbourhood 6 (odds ratio equal to 1.33). The main differences in the estimation of the neighbourhood effects were also observed in neighbourhoods 1 and 10 for the female population and in neighbourhoods 4, 5, 7, and 10 for the male population. In general, there were more discrepancies between the two estimation approaches for the effect of neighbourhood 10 than for the effects of the other neighbourhoods.

It is not possible to know exactly which of the two methods was more accurate in this setting, but, given the results of the simulation study, we can assume that the estimates based on the IPTW were more reliable because in the scenario closest to the real situation, this method performed better, with less bias.

7 Conclusions

The purpose of this work was to assess the performance of IPTW techniques for the estimation of causal effects in observational studies characterised by many treatments. The motivation for the study was the estimation of neighbourhood effects on the incidence of hospitalised fractures in the Italian city of Turin. One of the most intriguing methodological aspects of this study is that the number of treatments was large, and was thus not easy to handle. This was done by implementing simulation studies in which the IPTW approach was also compared to a standard logistic regression approach. Moreover, these approaches were applied to real data originating from our motivating case study.

The simulation study was performed under three possible scenarios for the allocation of individuals to neighbourhoods: one that was close to reality, one that had a complex misspecified treatment allocation, and one that had an extremely unbalanced initial situation. In all the scenarios, IPTW performed very well in terms of reducing the initial imbalance of the confounders across the different neighbourhoods. However, in the scenarios characterised by a higher initial imbalance, the bias of the estimated causal effect was higher than it was in the first scenario, which was characterised by a lower initial imbalance.

It is often stressed in the causal inference literature (**?**) that researchers should examine balance measures because a higher (residual) imbalance tends to be associated with more bias in the causal estimates. This was confirmed in our analyses, which showed that the bias tended to be higher for both the logistic regression and the IPTW approaches in scenarios characterised by higher initial (and residual) imbalances.

Table 4: Estimated neighbourhood effect on real data in terms of odds ratio (the reference is neighbourhood 6) without adjustment on observables (Crude), with the naive logistic regression (Logit) and with IPTW.

		Crude				**Logit**				**IPTW**			
Pop.	**Neigh.**	**Odds Ratio**	**Std.Err.**	**CI 95%**		**Odds Ratio**	**Std.Err.**	**CI 95%**		**Odds Ratio**	**Std.Err.**	**CI 95%**	
All	1	1.48	0.10	1.20	1.81	1.17	0.11	0.95	1.45	1.21	0.13	0.94	1.55
	2	1.31	0.10	1.09	1.59	1.34	0.10	1.11	1.62	1.33	0.11	1.07	1.64
	3	1.19	0.10	0.99	1.44	1.08	0.10	0.89	1.31	1.09	0.11	0.88	1.35
	4	1.30	0.10	1.07	1.59	1.15	0.10	0.94	1.41	1.15	0.11	0.92	1.43
	5	1.18	0.10	0.97	1.43	1.18	0.10	0.98	1.43	1.15	0.12	0.91	1.45
	7	1.46	0.10	1.19	1.78	1.33	0.10	1.09	1.63	1.33	0.11	1.06	1.66
	8	1.44	0.11	1.15	1.80	1.22	0.12	0.97	1.54	1.21	0.13	0.94	1.57
	9	1.29	0.10	1.05	1.58	1.25	0.11	1.02	1.54	1.25	0.12	1.00	1.58
	10	1.19	0.13	0.92	1.53	1.39	0.13	1.07	1.79	1.28	0.16	0.94	1.75
Women	1	1.49	0.12	1.19	1.87	1.22	0.12	0.97	1.55	1.32	0.14	1.00	1.75
	2	1.26	0.11	1.02	1.57	1.32	0.11	1.06	1.64	1.32	0.12	1.04	1.68
	3	1.23	0.11	1.00	1.52	1.14	0.11	0.92	1.41	1.18	0.12	0.93	1.49
	4	1.26	0.12	1.00	1.58	1.13	0.12	0.90	1.42	1.17	0.13	0.91	1.50
	5	1.11	0.11	0.89	1.38	1.13	0.11	0.91	1.41	1.14	0.14	0.87	1.48
	7	1.35	0.12	1.08	1.70	1.25	0.12	0.99	1.58	1.29	0.13	1.00	1.66
	8	1.45	0.13	1.13	1.86	1.27	0.13	0.98	1.64	1.27	0.15	0.95	1.69
	9	1.27	0.12	1.01	1.61	1.26	0.12	0.99	1.59	1.27	0.13	0.98	1.65
	10	1.26	0.15	0.95	1.68	1.54	0.15	1.15	2.05	1.43	0.17	1.02	2.00
Men	1	1.18	0.24	0.73	1.89	0.86	0.25	0.52	1.40	0.85	0.32	0.45	1.59
	2	1.46	0.20	0.99	2.18	1.40	0.20	0.95	2.10	1.37	0.24	0.86	2.18
	3	0.95	0.22	0.62	1.45	0.86	0.22	0.56	1.33	0.83	0.25	0.51	1.36
	4	1.32	0.22	0.86	2.03	1.17	0.22	0.76	1.81	1.07	0.25	0.65	1.75
	5	1.42	0.20	0.97	2.12	1.39	0.20	0.94	2.07	1.23	0.25	0.76	2.01
	7	1.72	0.21	1.15	2.62	1.59	0.21	1.05	2.42	1.50	0.24	0.93	2.42
	8	1.28	0.25	0.77	2.08	1.01	0.26	0.60	1.67	1.03	0.30	0.57	1.86
	9	1.31	0.22	0.85	2.02	1.23	0.22	0.80	1.92	1.23	0.26	0.74	2.04
	10	1.03	0.28	0.58	1.78	1.06	0.29	0.59	1.83	0.81	0.34	0.41	1.60

Our results indicate that IPTW is a promising approach for reducing the imbalance of confounders in a multi-treatment context, even in the presence of a number of treatments as high as 10.

However, the IPTW approach is more computationally demanding than a standard logistic regression (the computation of weights may last several hours if the number of treatments is high, as in our case). Future research may be devoted to investigating more computationally efficient approaches, which would be necessary if an even higher number of treatments than the 10 we considered here was used.

One limitation in the application we considered, but not in our simulations, is that the mobility of individuals between neighbourhoods may invalidate the SUTVA. Indeed, in some neighbourhoods, the individuals may have had a higher propensity to move and to be affected by other neighbourhoods. Since our focus was on older people, who tend to be a more stable population, this risk, in the empirical application, was limited. However, it would be interesting for future research to take this aspect into account as well.

References

Arpino, B. and Cannas, M. (2016). Propensity score matching with clustered data. an application to the estimation of the impact of caesarean section on the apgar score. *Statistics in medicine*, 35(12):2074–2091.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, 28(25):3083–3107.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424.

Austin, P. C. and Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (iptw) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in medicine*, 34(28):3661–3679.

Barnett, D. W., Barnett, A., Nathan, A., Van Cauwenberg, J., and Cerin, E. (2017). Built environmental correlates of older adultsťotal physical activity and walking: a systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*, 14(1):103.

Cannas, M. and Arpino, B. (2019). Comparison of machine learning algorithms and covariate balance measures in propensity score matching and weighting. *Biometrical Journal*, 61(4):1049–1072.

Cepeda, M. S., Boston, R., Farrar, J. T., and Strom, B. L. (2003). Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American journal of epidemiology*, 158(3):280–287.

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics*, 49(4):1231–1236.

Harding, D. J. (2003). Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology*, 109(3):676–719.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Lee, B. K., Lessler, J., and Stuart, E. A. (2011). Weight trimming and propensity score weighting. *PloS one*, 6(3).

Li, F., Zaslavsky, A. M., and Landrum, M. B. (2013). Propensity score weighting with multilevel data. *Statistics in medicine*, 32(19):3373–3387.

Linden, A., Uysal, S. D., Ryan, A., and Adams, J. L. (2016). Estimating causal effects for multivalued treatments: a comparison of approaches. *Statistics in medicine*, 35(4):534–552.

Lopez, M. J. and Gutman, R. (2017). Estimation of causal effects with multiple treatments: a review and new ideas. *Statistical Science*, 32(3):432–454.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., and Burgette, L. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in medicine*, 32(19):3388–3414.

McCaffrey, D. F., Ridgeway, G., and Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, 9(4).

Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social science & medicine*, 58(10):1929–1952.

Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L., and Griffin, B. A. (2006). Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the twang package. *Santa Monica, CA: RAND Corporation*.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, 82(398):387–394.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Roux, A. V. D., Borrell, L. N., Haan, M., Jackson, S. A., and Schultz, R. (2004). Neighbourhood environments and mortality in an elderly cohort: results from the cardiovascular health study. *Journal of Epidemiology & Community Health*, 58(11):917–923.

Sànchez-Riera, L., Wilson, N., Kamalaraj, N., Nolla, J. M., Kok, C., Li, Y., Macara, M., Norman, R., Chen, J. S., Smith, E., et al. (2010). Osteoporosis and fragility fractures. *Best practice & research Clinical rheumatology*, 24(6):793–810.

Setoguchi, S., Schneeweiss, S., Brookhart, M. A., and Glynn, R. J., . C. E. F. (2008). Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology and drug safety*, 17(6):546–555.

Tu, C. (2019). Comparison of various machine learning algorithms for estimating generalized propensity score. *Journal of Statistical Computation and Simulation*, 89(4):708–719.

Tu, C. and Koh, W. Y. (2016). Causal inference for average treatment effects of multiple treatments with non-normally distributed outcome variables. *Journal of Statistical Computation and Simulation*, 86(5):855–861.

Yen, I. H., Michael, Y. L., and Perdue, L. (2009). Neighborhood environment in studies of health of older adults: a systematic review. *American journal of preventive medicine*, 37(5):455–463.

A Descriptive statistics

Table 5: Descriptive statistics on the outcomes and the confounders by neighbourhoods.

	Neighbourhoods										
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
Incidence of Hospitalized Fractures (%)	1.05	0.93	0.85	0.92	0.84	0.71	1.03	1.02	0.92	0.85	0.90
Female (%)	60.60	57.40	58.88	59.50	56.75	56.67	58.67	58.97	57.48	55.07	58.63
Age (***Mean***)	71.99	70.63	71.22	71.35	70.48	70.43	71.14	71.68	70.87	70.02	70.96
Region of Birth(%)											
Piedmont	56.43	48.84	50.12	49.59	34.74	34.92	48.73	59.47	47.64	30.75	45.93
North of Italy	13.83	14.75	15.14	15.63	13.56	13.09	12.67	13.43	14.80	12.87	14.12
Center of Italy	3.74	3.51	2.76	2.89	2.59	2.73	2.57	3.24	3.24	2.42	2.97
South of Italy	21.19	27.54	26.73	27.04	39.98	41.91	31.25	19.59	27.42	47.62	30.93
Outside of Italy	4.81	5.37	5.25	4.86	9.13	7.36	4.77	4.28	6.91	6.33	6.05
Family composition (number of components) (%)											
Alone (1)	35.74	26.46	30.05	31.09	25.89	26.44	31.34	32.20	27.37	20.65	28.73
Married couple (2)	33.99	44.97	42.33	41.30	44.90	43.62	41.00	37.97	43.46	45.71	42.28
Married couple (> 3)	17.35	19.14	17.24	16.85	18.55	19.52	17.19	18.20	19.15	23.34	18.41
Not married couple (> 2)	12.92	9.42	10.38	10.76	10.66	10.42	10.47	11.63	10.02	10.29	10.58
Educational attainment (%)											
Primary or lower	26.05	40.73	43.04	43.15	60.99	61.42	48.37	31.42	47.19	63.03	46.94
Lower Secondary	25.73	34.15	32.40	31.43	29.22	28.84	30.53	28.47	33.88	27.95	30.69
Upper Secondary	25.43	18.38	17.28	17.80	7.64	7.23	14.03	22.96	13.70	6.73	14.94
Tertiary	22.79	6.74	7.28	7.61	2.15	2.51	7.08	17.16	5.23	2.29	7.42
Home owner (%)	71.15	81.43	77.99	75.48	71.21	72.54	76.87	78.99	80.01	79.48	76.34
Last observed professional condition (%)											
No observed work	13.75	11.61	13.28	14.11	14.25	15.72	16.65	13.92	14.74	12.08	14.00
Home-maker	34.05	35.24	36.02	34.81	35.02	33.74	33.51	34.85	33.50	36.53	34.74
Entrepreneur	16.90	5.75	6.73	7.07	2.45	2.34	6.09	13.63	4.89	1.73	6.34
White collars	24.45	26.73	24.53	24.67	17.32	17.14	22.79	24.75	23.29	14.88	22.33
Manual workers	10.85	20.66	19.44	19.33	30.96	31.06	20.96	12.85	23.59	34.78	22.59
Overcrowding (***Mean***)	0.64	0.74	0.78	0.77	0.84	0.82	0.78	0.66	0.79	0.76	0.77
Hypertension	51.16	54.89	54.49	54.73	57.04	58.85	56.52	53.09	56.45	58.08	55.58
Drugs (***Mean***)	7.35	7.84	7.72	7.78	8.09	8.16	7.87	7.61	7.96	8.18	7.86

B Parameters to simulate the three scenarios

Table 6: Parameters used to simulate the first scenario.

		Neighbourhoods								
Variables		**1**	**2**	**3**	**4**	**5**	**7**	**8**	**9**	**10**
(Intercept)	${}_1\beta_0$	-3.1640	-0.6430	-1.650	-2.1240	-0.0420	-1.6450	-2.7530	-1.4290	0.5070
Gender	${}_1\beta_1$	0.4420	0.1530	0.2410	0.2640	0.0140	0.1900	0.2980	0.1330	-0.1350
Lower Secondary Educ.	${}_1\beta_2$	0.7700	0.5770	0.5220	0.4930	0.0270	0.3320	0.6670	0.4550	-0.1300
Upper Secondary Educ.	${}_1\beta_3$	2.1240	1.3270	1.3030	1.3350	0.0760	0.9540	1.8080	0.9480	-0.2360
Tertiary Educ.	${}_1\beta_4$	3.0430	1.3690	1.4980	1.5460	-0.1280	1.3270	2.5410	1.0460	-0.2860
Hypertension	${}_1\beta_5$	-0.0910	-0.0820	-0.0820	-0.0910	-0.0770	-0.0200	-0.0910	-0.0570	-0.0390
Age	${}_1\beta_6$	0.0380	0.0100	0.0240	0.0250	0.0030	0.0190	0.0300	0.0140	-0.0130
Overcrowding	${}_1\beta_7$	-0.5490	-0.2530	0.0640	0.0370	0.0620	0.0020	-0.5560	0.0260	-0.4900
Drugs	${}_1\beta_8$	-0.0720	-0.0190	-0.0370	-0.0340	-0.0020	-0.0310	-0.0490	-0.0160	0.0100

Table 7: Parameters used to simulate the second scenario.

		Neighbourhoods								
Variables		**1**	**2**	**3**	**4**	**5**	**7**	**8**	**9**	**10**
(Intercept)	${}_2\beta_0$	2.8360	-1.0650	-0.6010	-0.8630	-0.9300	-0.2560	0.8170	-0.4910	-1.3490
Gender	${}_2\beta_1$	0.3390	1.2150	0.3480	-0.1640	0.1900	-0.1910	0.5620	0.4530	2.6710
Lower Secondary Educ.	${}_2\beta_2$	0.6120	0.4030	0.3750	0.3100	-0.1240	0.1800	0.3300	0.1930	-0.2680
Upper Secondary Educ.	${}_2\beta_3$	1.5400	0.9330	0.9470	0.9850	-0.2060	0.5550	1.1300	0.3670	-0.2740
Tertiary Educ.	${}_2\beta_4$	2.7400	1.3960	1.5980	1.5920	-0.3520	1.3990	2.0700	0.8550	0.5070
Hypertension	${}_2\beta_5$	-0.2960	-0.0730	0.5710	1.4020	0.1900	0.7710	0.2990	-0.5720	0.5760
Age	${}_2\beta_6$	-0.1010	0.0240	0.0060	-0.0080	0.0150	-0.0120	-0.0170	-0.0170	0.0470
Overcrowding	${}_2\beta_7$	-1.5120	-0.1240	-0.2110	-0.2790	0.3490	-0.3860	-2.1930	0.3660	-1.7120
Drugs	${}_2\beta_8$	-0.1870	-0.1570	-0.1780	-0.2260	-0.1110	-0.1630	-0.2530	-0.1020	-0.1700
Age^2	${}_2\beta_9$	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Overcrowding^2	${}_2\beta_{10}$	0.4830	-0.1700	0.0770	0.1970	-0.0520	0.2150	0.6720	-0.0840	0.5590
Drugs^2	${}_2\beta_{11}$	0.0190	0.0170	0.0080	0.0180	0.0050	0.0070	0.0140	0.0140	0.0180
Gender*Age	${}_2\beta_{12}$	0.0010	-0.0090	-0.0010	0.0110	0.0010	0.0060	-0.0040	-0.0030	-0.0390
Gender*Hypertension	${}_2\beta_{13}$	-0.0040	0.0800	0.0310	-0.0310	0.0800	0.1170	0.0530	0.0380	0.2030
Gender*Drugs	${}_2\beta_{14}$	0.0010	-0.0520	-0.0050	-0.0130	-0.0260	-0.0220	0.0040	0.0010	-0.0030
Age*Hypertension	${}_2\beta_{15}$	0.0080	0.0010	-0.0090	-0.0180	-0.0060	-0.0090	-0.0020	0.0080	-0.0130
Age*Drugs	${}_2\beta_{16}$	-0.0010	0.0000	0.0010	0.0000	0.0010	0.0010	0.0010	-0.0010	-0.0010
Hypertension * Drugs	${}_2\beta_{17}$	-0.0480	-0.0220	-0.0160	-0.0360	-0.0070	-0.0360	-0.0470	-0.0250	0.0170

Table 8: Parameters used to simulate the third scenario.

		Neighbourhoods								
Variables		**1**	**2**	**3**	**4**	**5**	**7**	**8**	**9**	**10**
(Intercept)	${}_3\beta_0$	-30.6410	-0.8200	-1.8790	-2.5810	-0.4830	-1.4070	-3.3250	-1.3680	-1.0100
Gender	${}_3\beta_1$	0.4420	0.1530	0.2410	1.3200	0.0140	0.1900	0.2980	0.0266	-0.1350
Lower Secondary Educ.	${}_3\beta_2$	0.7700	1.1540	0.5220	0.2465	0.0270	0.3320	0.6670	0.4550	-0.1300
Upper Secondary Educ.	${}_3\beta_3$	2.1240	1.3270	2.6060	1.3350	0.0760	0.4770	1.8080	0.9480	-0.2360
Tertiary Educ.	${}_3\beta_4$	3.0430	1.3690	1.4980	1.5460	-0.0640	1.3270	5.0820	1.0460	-0.2860
Hypertension	${}_3\beta_5$	-0.0910	-0.0820	-0.0082	-0.0910	-0.0770	-0.2000	-0.0910	-0.0570	-0.0390
Age	${}_3\beta_6$	0.3800	0.0100	0.0240	0.0250	0.0030	0.0190	0.0300	0.0140	-0.0013
Overcrowding	${}_3\beta_7$	-0.5490	-0.2530	0.0640	0.0370	0.6200	0.0020	-0.0556	0.0260	-0.4900
Drugs	${}_3\beta_8$	-0.0072	-0.0190	-0.0370	-0.0340	-0.0020	-0.0310	-0.0490	-0.0160	0.1000

C Parameters to simulate the outcome

Table 9: Parameters used to simulate the outcome.

Variables	**Parameter**	**Value**
(Intercept)	β_0	-13.553
Gender	β_1	0.740
Lower Secondary Educ.	β_2	-0.012
Upper Secondary Educ.	β_3	0.098
Tertiary Educ.	β_4	0.128
Hypertension	β_5	0.029
Age	β_6	0.100
Overcrowding	β_7	0.006
Drugs	β_8	0.077
Neighbourhood 1	β_9	0.820
Neighbourhood 2	β_{10}	1.310
Neighbourhood 3	β_{11}	0.375
Neighbourhood 4	β_{12}	0.720
Neighbourhood 5	β_{13}	0.915
Neighbourhood 7	β_{14}	1.430
Neighbourhood 8	β_{15}	0.950
Neighbourhood 9	β_{16}	1.020
Neighbourhood 10	β_{17}	1.535

D R code for the simulation study

In this Section we report the R code used for the simulations described in section 4. As far as was possible, we used the same notation (variables names, parameters name, etc.) as in the main text.

```
###In this code X1 represents the gender, X2 the age, X3 the
educational attainment, X4 the overcrowding, X5 the hypertension
and X6 the drugs prescription. Moreover, for the second scenario,
quadratic transformation of age (X7), overcrowding (X8) and drugs
prescriptions (X9) have been created.

###Create the counfounders' matrix with respect to the first and
the third scenarios.

X<-model.matrix(~X1+X2+X3+X4+X5+X6)

###Create the counfounders' matrix with respect to the second
scenario.
X<-model.matrix(~X3+X4+X7+X8+X9+(X1+X2+X5+X6)^2)


###Create probability vectors to live in every neighbourhood
(treatments), chosing the right X
and beta matrices according to the considered scenario.

denominator<-c(rep(1, dim(sample)[1]))
p<-matrix(data=NA, nrow = dim(sample)[1], ncol=9)
for (i in c(1:9) ){
p[,i]<-exp(X%*%beta[,i])
denominator<-denominator+p[,i]
}
p<-cbind(p[,1:5],c(rep(1, dim(sample)[1])),p[,6:9])
prob.sim<-p/denominator

###Once the number, n, of desired replicates has been set it is
possible to proceed with the simulations.
for (j in 1:n){

###Treatment generation
assign.treat = t(apply(prob.sim, 1, rmultinom, n = 1, size = 1))
s1 = cbind.data.frame(sample, treat_sim = apply(assign.treat, 1,
function(x) which(x==1)))

###Create probability vectors with respect to the occurence of
```

```
the outcome
XO<-cbind(X, assign.treat[,c(1:5,7:10)])
sim.prob<-(1+exp(-(XO%*%beta_out)))^(-1)

###Outcome generation
unif<-runif(dim(sample)[1],0,1)
s1$sim.y<-ifelse(sim.prob>unif,1,0)

###Computation of weights with the twang package
mnps_s1 <- mnps(as.factor(treat_sim) ~ X1+X2+X3+X4+X5+X6,
data = s1,
estimand = "ATE",
verbose = FALSE,
stop.method ="es.max",
n.trees = 3000)

###Saving balance measures
for (b in 1:10){
row<-(j-1)*10+b
bal<-bal.table(mnps_s1$psList[[b]])
balance[row,]<-bal$es.max.ATE[,5]
balance_now[row,]<-bal$unw[,5]
}

###Estimation of the treatment effect
s1$t.sim<-relevel(as.factor(s1$treat_sim), ref=6)
s1$weight<-get.weights(mnps_s1)
weights[,j]<-s1$weight
design.s1<-svydesign(ids=~1, weights=~weight, data=s1)

####With IPTW and no trimming
glm.s1<-svyglm(sim.y~t.sim, design=design.s1, family=quasibinomial())
mccaffrey[j,]<-glm.s1$coefficients[-1]
ci<-confint.default(glm.s1, 2:10)
for (c in 1:9){ ci_mcc[j,c]<-between(true[c], ci[c,1], ci[c,2])}

####With logistic regression model
glm.s1_now<-glm(sim.y~t.sim+X1+X2+X3+X4+X5+X6, data=s1,
family=binomial())
logistic[j,]<-glm.s1_now$coefficients[2:10]
ci<-confint.default(glm.s1_now, 2:10)
for (c in 1:9){ ci_log[j,c]<-between(true[c], ci[c,1], ci[c,2])}

#####Create trimmed weights
percentiles<-c(c(75,85:99)/100)
thresholds<-quantile(s1$weight, percentiles)
```

```
w_trim<-matrix(data=NA, nrow=dim(s1)[1], ncol=length(percentiles))
for(i in 1:length(thresholds)){
w_trim[,i]<-ifelse(s1$weight>thresholds[i],thresholds[i],s1$weight)
}
trimmed<-cbind(s1,w_trim)

####With IPTW and different levels of trimming, listed in
the vector percentiles. For every level of trimming a different
column in the matrix trimmed has to be considered as in the
following example:

design.t<-svydesign(ids=~1, weights=~‘1‘, data=trimmed)
glm.t<-svyglm(sim.y~t.sim, design=design.t, family=quasibinomial(),
maxit=100)
trimming75[j,]<-glm.t$coefficients[2:10]
ci<-confint.default(glm.t, 2:10)
for (c in 1:9){ ci_t75[j,c]<-between(true[c], ci[c,1], ci[c,2])}
```

Acknowledgements

The authors thank Prof. Giuseppe Costa and his collaborators of the Unit 'SCaDU Servizio Sovrazonale di Epidemiologia' in Grugliasco (Turin, Italy) for their useful suggestions and support in the data management. The data used for the research are part of the health administrative datasets of Turin, Italy. They have been managed following a formal agreement between the SCaDU Service and the Department of Statistical Science of the University of Padova.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it
Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy