Health Programme Evaluation by Propensity Score Matching: Accounting for Treatment Intensity and Health Externalities with an Application to Brazil

by

Rodrigo Moreno-Serra[a]

[a] Department of Economics & Related Studies, University of York

November 2008

Abstract

Most of the literature on health programme evaluation has estimated average programme impacts relying on either: (*i*) data on the presence or absence of an intervention in a particular locality, or (*ii*) data on individual participation in the health programme. By estimating an average health impact which is independent of the programme's population coverage, the empirical approaches of these studies overlook the important fact that public health interventions create externalities whose magnitude depends crucially on the number of covered individuals in a locality. The main contributions of this paper are to suggest and apply an empirical approach for the impact evaluation of public health interventions which also takes into account treatment externalities, when non-experimental, routine data are available and under the assumption of selection on observables. The proposed framework involves the computation of average treatment effects by a propensity score matching-difference-in-differences estimator adapted to the case of multiple treatments, jointly evaluating the impact of different programme coverage levels. The methods are used to conduct an impact evaluation of the Family Health Programme (*Programa Saude da Familia*—PSF), the broadest health programme ever launched in Brazil, on adult and child health.

Contact author: Rodrigo Moreno-Serra, Centre for Health Economics, Alcuin A Block, University of York, York, YO10 5DD, United Kingdom. Tel. (1904) 321 411. Email: rams500@york.ac.uk

Keywords: Health programme evaluation; multiple treatments; propensity score matching; Brazil.

1. Introduction

In general terms, impact evaluations of a health programme aim to answer the fundamental counterfactual question: how would the health conditions of treated individuals have evolved in the absence of the programme? Or, analogously, how would those who were not exposed to the programme have fared in the presence of it? Difficulties in answering such a question rise immediately, as at a given point in time individuals are observed in only one situation, either exposed or not exposed to the programme. As many aspects may have varied from the time individuals were exposed to the intervention, it is usual to measure the programme's average impact on a group of individuals by comparing the evolution of some indicators in this group with the evolution of the same indicators in a similar group of individuals not covered by the programme. However, individuals exposed to a programme are usually different in a set of unobservable or unobserved characteristics—such as initial health status and health risk aversion—from those individuals who are not covered by the intervention, making it difficult to isolate the differences between both groups which are due to already existing distinctions before treatment (the selection bias) from those which are due solely to the programme's impact. In short, the major problem relies on constructing an adequate comparison group.

An evaluation design in which the selection bias problem tends to disappear is that in which treatment and comparison groups are randomly selected from a large population of potential beneficiaries, such as individuals or localities. In this case, if the randomisation of treatment assignment has been adequately performed, it can be assured that any statistically significant difference in health outcomes between both groups is due solely to the programme's impact.[1] In most situations, nevertheless, health programmes have been purposively implemented (for instance, by targeting individuals or areas with worse than average health status) and/or require individuals to self-select into the programme by taking up the benefits. And if all the researcher has for evaluating these interventions is non-experimental data, explicitly dealing with the potential bias caused by omitted variables—either unobserved or intrinsically unobservable—is of crucial importance for the reliability of the estimates of the programme's impact.

In many cases, health interventions present another important characteristic that should be taken into account when their impacts are to be measured. Most of the theoretical and empirical literature on programme evaluation relies, at least implicitly, on the so-called "stable unit treatment value assumption" (SUTVA), which encompasses two components: in the first place, all treated individuals are assumed to receive the same active treatment and all comparison individuals are assumed to get the same comparison treatment; the second component is the assumed absence of interference between units, in the sense that the values of treated and untreated outcomes for a given individual are not influenced by the treatment status of other individuals. Although the validity of the two SUTVA components might be questioned in specific settings, the second aspect mentioned above may be particularly unrealistic in the context of health programmes. Treatment benefits from public health interventions usually positively affect untreated individuals as well, such as in the classical examples of immunisation campaigns and programmes aimed at

[1] Randomised studies for the evaluation of social programmes have other noteworthy drawbacks though, a topic discussed in detail by Heckman and Smith (1995).

reducing the prevalence of communicable diseases. These treatment externalities pose a significant challenge to the assessment of a programme's impact through non-experimental or individually randomised studies, since there is the possibility of non-negligible treatment benefits accruing to the comparison group. This would lead to an underestimation of the *total* programme effects when comparing the average outcomes of treatment and comparison samples, as clearly demonstrated by Miguel and Kremer (2004).

Miguel and Kremer (2004) also demonstrate (within an experimental setting) that it is sometimes possible to alleviate deviations from SUTVA through design; for example, by considering higher-level randomisation units rather than individuals. Non-experimental evaluations of health programme treatment effects can deal with deviations from SUTVA in a similar way, by considering the availability of a health programme in a given geographic area as the treatment variable of interest. For example, consider the all too common situation in public health policy in which a health authority implements an intervention in some geographic areas selected according to pre-specified criteria. In the treated areas—those where the health intervention has been implemented—it is not necessarily the case that all residents actually receive the intervention: there might exist a prioritisation of implementation across sub-areas within those treated areas according to observed health needs, for instance, due to budgetary constraints that preclude universal coverage (as in the PSF case). A treated individual can be defined here as one who resides in an area where the programme is available. For simplicity, consider now only two areas. If the treated area is far enough from the untreated area (the one where the health intervention has not been implemented) so as to preclude spillovers from occurring between localities, the SUTVA component of "no interference between units" is more likely to be valid and it is possible for the estimated average treatment effects to take into account treatment externalities accruing to people living in the treated area but who have not directly received the intervention.[2]

Nonetheless, even when impact evaluation studies have adopted an empirical strategy in the spirit of the one described above for estimating average treatment effects, a simple indicator variable (i.e. a dummy) has normally been used to represent the presence or not of the relevant intervention in a given area as the treatment of interest.[3] The empirical approach of these studies overlooks the important fact that the magnitude of any programme-related health externalities within a locality is likely to depend crucially on the number of *actually treated* individuals in the same locality.[4] There is a fundamental identification problem arising from the fact that only one mean impact is estimated which is irrespective of the number of individuals who actually receive the programme's services in a given locality, an important dimension of the *intensity* of treatment.

Yet in several contexts where health programmes have been implemented in a phased manner and with different population coverage levels across areas, evaluation research can in fact take treatment intensity into account with non-experimental data by using a measure of the programme's population coverage as the treatment variable of interest. This paper suggests an empirical framework that involves the computation

---

[2] This is the basic definition of an "intention-to-treat" estimator.

[3] See Angeles et al. (2005), Armecin et al. (2006), Attanasio and Vera-Hernandez (2004), Frankenberg et al. (2005) and Jalan and Ravallion (2003), just to cite a few recent examples.

[4] As also found by Miguel and Kremer (2004).

of a number of average treatment effects through comparisons between the health impacts of alternative “programmes”, where a specific number of different coverage levels play the role of the compared alternatives, thus allowing the researcher to investigate the effect of different treatment intensities on individual health outcomes. This generally applicable empirical strategy is used to perform an impact evaluation of the Brazilian Family Health Programme (*Programa Saude da Familia*—PSF) on the health outcomes of adults and children living in regions with different programme coverage levels, based on a propensity score matching-difference-in-differences estimator adapted for the multiple treatments setting and data from two repeated cross-sections. In addition to being one of the few econometric evaluations of PSF impacts, this paper presents an empirical approach which has the advantage of incorporating into each estimated treatment effect—but not separately quantifying—the (possibly non-linear) treatment externalities arising from the level of population coverage of a particular health intervention.

The paper is organised as follows. Section 2 offers a brief description of the Family Health Programme and its institutional context, the Brazilian health system. The suggested general empirical approach to comprehensively evaluate health programme impacts is outlined in Section 3, whereas Section 4 describes the data used in the estimations. The results of the specification tests and estimations for children and adults are presented in Section 5. Section 6 discusses the empirical results and concludes.

2. The Brazilian health system and the Family Health Programme (PSF)

The Brazilian national health system (*Sistema Unico de Saude*—SUS) is based on three main general principles. Firstly, access to health care must be universal and provided free of charge at the point of use to all individuals, i.e. on the basis of need rather than ability to pay. Secondly, free health care must be provided at all levels of complexity, from preventive actions to the most complex forms of hospital treatment. Finally, the responsibility for the funding and the actual provision of health care actions is to be shared between the three government tiers—federal (national), state and municipality—with an increasing emphasis on managerial decentralisation towards the municipality level since the inception of SUS in 1988. The financial resources for funding the health care sector are collected by the federal government through general taxes and then transferred to states and municipalities. States usually receive the bulk of their transfers relative to the provision of hospital services whereas municipalities, in addition to their general managerial and coordination responsibilities regarding the provision of health services at all levels of complexity in the locality, are normally directly responsible for the provision of primary care services. With this aim, municipalities receive their share of the total health budget according to a formula which includes a fixed component (a per capita amount) and a variable component for those municipalities implementing so-called “strategic actions”; these are usually health programmes of a preventive nature but also include other initiatives such as the provision of medicines funded by the public system.

Among the strategic actions mentioned above, one of the most important is the Family Health Programme (*Programa Saude da Familia*—PSF). This programme is a federal initiative officially launched in 1994 by the Ministry of Health, though a more restricted version of the PSF, known as the Community Health Agents Programme, had been in place mainly in rural areas since June 1991. The first PSF teams were formed in January 1994 with the aim of performing preventive and health promotion activities for all the individuals in a family, in a global and continued manner (Ministerio da Saude, 2001). As such, PSF is an integral part of a broader federal strategy for the health sector which seeks the substitution of a model based on curative care towards a focus on primary care activities. It is centred on the Family Health Unit, a public health unit that provides the physical infrastructure for the work of the family health teams; since the PSF is intended to be an instrument within a wider reorganisation of priorities for the health sector, the implementation of the programme in a given locality is expected to take advantage of the existing infrastructure and therefore does not normally lead to the creation of new health facilities, except in the case of the municipalities without any basic health infrastructure (these tend to be also the poorest municipalities, located chiefly in rural areas). The Family Health Unit should be able to monitor health needs and provide primary care services for the population living in a specific area within the municipality, and refer those individuals to higher levels of health care complexity when necessary.

The officially stated goal for the PSF is fairly broad—namely, "to improve the health conditions of the covered families" (Ministerio da Saude, 2001)—and so are the profile of covered individuals (male and female adults, seniors, children) and the health actions performed accordingly to these different profiles. PSF services are provided by multi-professional work teams known as family health teams (FHT), which are formed at least by a full-time generalist doctor or a family doctor, a nurse, an assistant nurse, and four to six community health agents. Although other health professionals (such as dentists and psychologists) can in principle be incorporated if deemed appropriate based on local needs and possibilities, the basic structure for a FHT described above must always be present for the municipality to be eligible to receive the federal transfers and incentives corresponding to the PSF (these are explained below). According to the guidelines developed by the Ministry of Health, each FHT should cover at most 4,500 individuals and the municipality itself must set the number of community health agents depending on the actual number of individuals covered by a FHT, yet each agent should not be responsible for more than 750 people (or 150 families).

The family doctor represents the highest level of health care attention within the PSF and is responsible for offering primary care services and referring individuals to secondary and tertiary care. The nurse should supervise the work of the assistant nurse and community health agents, in addition to performing primary care activities at the Family Health Unit or the person's home. Important as family doctors and nurses are for the PSF, the community health agents represent the vital core of the family health strategy. These professionals play the role of a bridge between families and health services and are in fact supposed to be the first contact point of the former with the latter. Community health agents must visit each household under their responsibility at least once a month, as well as map each area, register the families, stimulate healthy lifestyles and perform preventive health actions. Those agents are recruited among individuals who have been living in the covered locality for at least two years, thus being in an advantageous position for gaining the residents' trust, knowing their real health conditions and identifying the locality's priority areas for intervention.

From a practical perspective, community health agents are able to offer the most basic health services related to prevention and health promotion including, among other activities: the regular monitoring of the children's vaccination schedule (referring the child to a health centre in case they are behind schedule) and the weight of children aged less than two years old (helping the premature detection of nutritional deficiencies); promotion of the use of oral rehydration therapy to treat children affected by diarrhoeal diseases; identification of pregnancy cases in the families, referring expecting mothers to pre-natal care, following up on the frequency of such consultations and advising on the importance of breastfeeding and adequate immunisation; provision of information to women about the risks and importance of preventive exams against breast and cervical cancer, and encouraging regular examinations; provision of information about family planning methods and preventive actions against sexually transmissible diseases; and the monitoring of the blood pressure of individuals affected by hypertension as well as raising awareness about the risks and control of hypertension and diabetes.

Although all three government levels hold responsibilities regarding the adequate functioning of the programme (including its financing), the main features concerning the PSF's population coverage in a given area constitute basically a political decision made by the municipality. This process can be separated into two steps: firstly, government officials of a given municipality decide whether the programme will be implemented there at all; secondly, if implementation goes ahead, the local government determines the programme's coverage by specifying the number of FHTs that will be formed and which sub-areas (usually neighbourhoods) within the municipality will be given priority regarding their allocation. Also important, individuals do not self-select into the PSF as it usually occurs with other interventions—at the individual level, PSF works as a mandatory programme. Instead, *municipalities* are the units that "self-select" into the PSF. After a municipality opts for adopting the programme and decides on the areas that will receive its services, all individuals living in any given covered area are to be registered and visited by FHTs. In general, most municipalities have placed their FHTs firstly in the poorest and unhealthiest neighbourhoods, normally using simple indicators to guide such prioritisation of areas—e.g., average income, Human Development Index or infant mortality levels (see, for instance, Ministerio da Saude, 2005).

In practice, municipalities are also in charge of most decisions concerning the management of the programme once it has been implemented. Municipalities should set up the Family Health Units, integrating them into the local health infrastructure and establishing rational links with the higher levels of care complexity in the health system; they are also responsible for hiring the health professionals required and for paying current and capital expenditures associated with the programme. On the other hand, the federal level is responsible for the definition of norms and guidelines concerning the programme's implementation and, jointly with the states, the provision of technical support related to the adoption, definition of strategic priorities and management of the programme. In spite of the very general nature of the programme, Ministry of Health guidelines specifically encourage FHTs to be trained and perform actions with a focus on the following main areas: child health, health during pregnancy, hypertension monitoring, diabetes, tuberculosis and Hansen's disease (Ministerio da Saude, 2004).

The federal government and the states also play an important part as far as the funding of PSF activities is concerned through their transfers to municipalities. Even though the latter enjoy a considerable degree of autonomy on managing and expanding the programme, local administrations must be willing to follow the federal guidelines on the family health teams' basic composition and activities in order to qualify for the corresponding financial transfers. Until 1997, municipalities received block transfers earmarked to health care from the federal government with no attached criteria for the allocation of resources between primary and other levels of care. A new mechanism for such transfers was implemented from 1998 onwards, explicitly assigning the amount of monies accruing to primary care and establishing additional financial incentives for the implementation of health actions considered strategic by the national government. In this context, since 1998, the amount of corresponding federal resources transferred to a given municipality after the adoption of PSF directly depends on the population coverage achieved in the locality: higher coverage levels correspond to a larger amount of annual transfers per family health team. In 2001, municipalities received R$28,000 per year per FHT (around US$12,000 in 2001 prices) for a total coverage level below 5%, reaching R$54,000/year (US$23,300) per FHT if the total coverage level was higher than 70%. In addition to these monies, municipalities received a one-off, lump-sum payment of R$10,000 (US$4,300) for every newly formed FHT (Ministerio da Saude, 2001). Finally, since there are no explicit rules regarding the format and magnitude of the financial support from states to municipalities, some states (e.g., Parana) limited their regular support to the donation of physical inputs such as medical equipment, whereas other states like Ceara, Minas Gerais and Sao Paulo set up financial transfer schemes similar to the federal one (Marques and Mendes, 2002).

The PSF is the broadest health programme ever launched in Brazil, with an ever-increasing population coverage at the national level which reached more than 80 million people in 2006 and a large and growing amount of public resources invested in it (around 10% of the total federal health spending in the same year and over a quarter of the total federal transfers to primary care). Yet there is an almost complete dearth of evidence concerning the true effects of the programme's coverage on population health status, with at best some preliminary findings at the aggregate level that a higher coverage level of the programme in a given area is associated with decreased infant mortality rates (Moreno-Serra, 2005; Macinko et al., 2006). In the next section, I describe and justify an empirical approach for evaluating the impact of the PSF on individual health status, taking into account any programme-related health externalities in the estimated treatment effects. This approach can in principle be generalised to evaluate other health interventions that, like the PSF, are implemented with varying degrees of coverage across a given geographic area, using publicly available, routine data.

3. Empirical strategy

Borrowing Blundell and Costa-Dias (2000) criteria, the plausibility of an estimator to evaluate the impact of a particular programme must be assessed based on (*i*) the treatment effect of interest (an average treatment effect concerning only the treated or the general population, for instance); (*ii*) the programme's institutional characteristics; and (*iii*) the nature of the data available. The availability of two cross-sections of data on individual health outcomes and PSF population coverage across Brazilian metropolitan regions (as detailed in the Data section below) allows the present study to account for the fact that the magnitude of PSF's overall externalities depends on the programme's coverage level, that is, the percentage of residents actually treated within a given region. Higher levels of PSF coverage are likely to decrease the spread of communicable diseases and to increase the probability of a given resident interacting with covered people, helping disseminate good health practices and thus also potentially increasing the intensity of any programme-related health externalities.

I implement a general empirical approach along the lines described above by using propensity score matching estimators adapted to the case of multivariate discrete treatments, proposed almost simultaneously by Imbens (2000) and Lechner (2000), coupled with a difference-in-differences approach. The use of propensity score matching estimators coupled with difference-in-differences has now become standard in the evaluation literature for the case of a single treatment or intervention, though not so in the context of multiple treatments evaluated simultaneously as in this paper. The main advantage of such estimators relative to alternative methods used in the presence of non-experimental data relies on their well-known semi-parametric nature, allowing the estimation of treatment effects without imposing restrictive distributional assumptions to the data generating process. However, as I describe below, these estimators do rely on other important assumptions.

Average treatment effects with multiple treatments: definition and identification

Applying the definitions introduced by Lechner (2000) to the more specific context of this paper, let a given health programme be implemented in a group of localities according to sequentially increasing, mutually exclusive coverage levels denoted by $l \in \{0,1,2,...,L\}$. A given individual *i* who lives in a locality with a coverage level *l* will have then only one element of the health outcomes set $\{Y^0, Y^1, Y^2, ..., Y^L\}$ observed at any given point in time, the remaining being her counterfactual outcomes. The treatment variable *D* can thus assume one of $(L+1)$ discrete values: $D \in \{0,1,2,...,L\}$.

The average treatment effects usually defined in the impact evaluation literature for the single treatment case are expanded so as to encompass the presence of multiple treatments, although the focus remains on pairwise comparisons between the health effects of two different coverage levels, say $l_0 = 0$ and $l_1 = l$, $l_1 > l_0$.[5] The causal effects of interest are now related to the difference $Y^l - Y^0$, that is, the effect of being exposed to treatment level *l* and not being exposed to treatment level 0. As shown by Lechner (2000), a number of average treatment effects can then be defined; in particular, the *average treatment effect on the treated (ATT)*—the average programme's impact among those who reside in a locality with a coverage level *l* when compared to those who live in a locality with coverage level 0, can be defined as:

$$ATT^{l,0} = E\left[Y^l - Y^0 \mid D = l\right] = E\left[Y^l \mid D = l\right] - E\left[Y^0 \mid D = l\right] \tag{1}$$

Hence, in the context of a health programme such as the PSF, the ATT is equivalent to the marginal gain (in terms of health outcomes) accruing to a randomly selected individual from a locality with coverage level *l*, relative to what would have been her outcome if she lived in a locality with coverage level 0. As in the single treatment case, the ATT can be consistently estimated using propensity score matching methods in a multiple treatment setting if two fundamental assumptions about the treatment—or, in this case, the PSF coverage level to which an individual is exposed—hold: *weak unconfoundedness* and *overlap*. Let the ATT of interest be that associated to increasing the PSF coverage level from 0 to *l*, and let $p^0(X)$ and $p^l(X)$ be the individual probabilities of being exposed to coverage levels 0 and *l*, respectively, given a vector

[5] Note that the coverage level l_0 needs not be "zero coverage", representing instead any coverage level (zero or positive) chosen as the comparator.

of observed individual covariates *X*. The two ATT identification assumptions can then be expressed as:[6]

ASSUMPTION 1 (WEAK UNCONFOUNDEDNESS FOR MULTIPLE TREATMENTS)

$$Y^0 \perp D \mid X = x,\ D \in \{0, l\} \Rightarrow Y^0 \perp D \mid p^{l|0,l}(X) = p^{l|0,l}(x),\ D \in \{0, l\} \tag{2}$$

where $p^{l|0,l}(x) = \Pr(D = l \mid D \in \{0, l\},\ X = x) = \dfrac{p^l(x)}{p^0(x) + p^l(x)}$ is the *generalised propensity score*.[7]

ASSUMPTION 2 (OVERLAP FOR MULTIPLE TREATMENTS)

$$p^l(x) < 1 \tag{3}$$

Assumption 1 states that individual exposure to coverage level 0 or *l* of the programme is independent of the potential health outcome under coverage level 0 if the relevant observable covariates (i.e. those that jointly affect the potential outcomes and coverage level's exposure) are controlled for. Unobserved characteristics will only lead to selection bias if they are correlated both with exposure to a given PSF coverage level and potential health outcomes, for instance if "more health-concerned" individuals are also more likely to migrate to areas where the programme's coverage level is higher in order to gain access to it, and this selective migration is not observed by the researcher. Importantly, if weak unconfoundedness holds by conditioning on *X*, all biases due to observable characteristics are also removed by conditioning solely on a scalar representing the individuals' conditional probability of exposure to the coverage level of interest given the set of observable pre-treatment characteristics *X*, the generalised propensity score, and hence the weak unconfoundedness assumption remains valid.

Assumption 2 states that there is overlap between "treatment" and "comparison" samples (individuals exposed to coverage levels *l* and 0, respectively) at all values of *X* observed in the "treatment" sample. This assumption refers to the joint distribution of the treatment variable and covariates, implying that, conditional on *X*, there must be other variables which affect exposure to the alternative programme's coverage levels. If the weak unconfoundedness assumption also holds, these unobserved variables are not correlated with the potential health outcomes and the counterfactual $E\left[Y^0 \mid D = l\right]$ can be consistently estimated as $E\left[E\left[Y^0 \mid p^{l|0,l}(X),\ D = 0\right] \mid D = l\right]$.[8]

With data from two repeated cross-sections—before and after the intervention was put in place for the first time—and individual information on exposure to a coverage level 0 representing the situation of absence of the programme over time, a simple difference-in-differences (DD) approach could obviously be used instead of

[6] The formal proofs can be found in Lechner (2000).

[7] Note that this case is similar to that of a binary treatment variable for which $p^0(x) + p^l(x) = 1$, but recall that, in the general case of multiple treatments, $p^0(x) + p^l(x) < 1$.

[8] Thus, a sample reduction property is derived in the multiple treatments setting. If the interest lies in estimating the ATT for a particular pairwise comparison of treatment levels, weak unconfoundedness can be assumed to hold only for the sub-sample of individuals subjected to the compared treatment levels and this sub-sample is the only one required for the empirical analysis. Moreover, a conditioning set reduction is achieved whereby propensity score matching can be based on the single dimension conditioning set $p^{l|0,l}(X)$, a composite individual index. See Lechner (2000) and Imbens (2000).

propensity score matching to assess the health impacts of being exposed to coverage level l compared to no coverage. This approach can be implemented through a regression framework of the type:

$$Y_{it} = l_i + \delta_A + \beta\left(l_i * \delta_A\right) + \gamma X_{it} + \varepsilon_{it} \tag{4}$$

where Y_{it} is the health outcome of interest for individual i measured at time t, l_i is an indicator for whether the individual lives in the treatment region l, δ_A is an indicator for whether the individual is being observed at the period after programme implementation and X_{it} is a vector of individual covariates thought to potentially influence both individual exposure to the programme and the health outcome. The DD estimate of the ATT of being exposed to coverage level l instead of not being covered by the programme is then given by the pooled ordinary least-squares estimate of the coefficient β associated with the interaction between living in the treatment region l and being observed after the programme's implementation.

However, the availability of repeated cross-sections allows the researcher to employ a potentially more robust empirical strategy for estimating the ATT of being exposed to a given coverage level of the programme. In this context, it is possible to combine a DD estimator with a propensity score matching procedure to construct the required counterfactuals, so as to compare the change in health outcomes for individuals living in an area with coverage level l (the treatment area) to the change in health outcomes for similar individuals living in the area with coverage level 0 (the comparison area), where the change is measured relative to the pre-programme benchmark—that is, health outcomes before the programme was implemented. The ATT of exposure to PSF's coverage level l on individuals residing in this treatment area, compared to the absence of the programme (exposure to coverage level 0) can then be estimated as (Blundell and Costa-Dias, 2000):

$$\hat{\beta}_{PSDD} = \frac{1}{N^*_{l_A}} \sum_{i\in\{l_A\cap S^*\}} \left[\left(Y^l_{i,A} - \sum_{j\in\{0_A\cap S^*\}} W_{ij} Y^0_{j,A} \right) - \left(\sum_{j\in\{l_B\cap S^*\}} W_{ij} Y^l_{j,B} - \sum_{j\in\{0_B\cap S^*\}} W_{ij} Y^0_{j,B} \right) \right] \tag{5}$$

In the above definition of the ATT estimator of propensity score matching with difference-in-differences (PSDD) using repeated cross-sections, l_B, l_A, 0_B and 0_A stand for the treatment and comparison areas before and after the programme, respectively; S^* is the joint common support (the subset of individuals living in the treatment area after the programme who are matched for the construction of each and every counterfactual above, which depends on the particular matching method used) and $N^*_{l_A}$ represents the subset of individuals living in the treatment area after exposure to the programme and who belong to the joint common support. Finally, Y is the individual health outcome of interest and W_{ij} is the weight attributed to matched individual j when compared to "treated" individual i (which also depends on the matching method chosen). As it is clear from (5), in this PSDD framework with repeated cross-sections matching has to be performed three times for each individual living in the treatment area: to find comparable individuals living in the treatment area prior to the programme and comparable individuals living in the comparison area pre- and post-programme. Furthermore, the chosen comparison coverage level must be zero or sufficiently close to zero, since pre-programme data are only informative about potential health outcomes in the absence of the intervention.

The main appeal of the PSDD approach described above comes from the possibility of combining the strengths of the semi-parametric propensity score matching and difference-in-differences methods. In addition to its semi-parametric nature, matching procedures ensure that a given individual living in the treatment region of interest is compared, in terms of health outcomes, only to her counterparts in the comparison area who are similar in observable characteristics (with the outcomes of the comparison individuals weighted according to how close they are from the treated individual in terms of observables) and, unlike an OLS procedure, does not “force” the data by extrapolating results outside the region of common support. Coupling a propensity score matching procedure—which is only able to deal with observable confounders—with a DD approach offers the scope for representing an unobserved determinant of individual exposure to a given PSF coverage level, decomposed into group and time-specific components of the error term (Blundell and Costa-Dias, 2000; Smith and Todd, 2005). Once the three counterfactuals in (5) have been constructed by a selected matching procedure, the ATT of interest is estimated as under the additional assumptions of separable additivity of the group and time effects.

Due to its aforementioned desirable characteristics, the PSDD approach is given preference over the simple DD estimator (4) in the empirical application below. The main institutional features of the PSF also suggest that the PSDD estimator is well-equipped for the proposed impact evaluation task. For this case in particular, individual self-selection into the programme seems to be less of a problem: as previously described, PSF works as a mandatory programme for individuals living in areas covered by it, who will necessarily be visited by Family Health Teams; additionally, all residents of a given region will be “mandatorily” exposed to the PSF coverage level observed in that region and to any health externalities arising from residents actually visited by the PSF teams there.[9] Thus, being treated by the programme is arguably exogenous from the point of view of the individual after matching on the relevant observables is performed, and an impact evaluation of the PSF which uses a matching approach in the ATT estimations—and exposure to the programme coverage at the region level as the treatment variable, in order to capture health externalities—is considerably less likely to suffer from the problem of individual self-selection into treatment which plagues a good amount of the programme evaluation literature.

However, if weak unconfoundedness (Assumption 1) does not hold in the data even after controlling for the available observable characteristics of individuals through matching, the PSDD approach can still provide an unbiased estimate of the ATT of interest provided that the unobserved factors influencing both potential health outcomes and exposure to a given PSF coverage level are time-invariant (at least during the study period). This is equivalent to imposing the identifying assumption of

[9] It might be argued that individuals can “opt-out” of the programme by refusing the access of PSF professionals to their homes. Although this is of course a possibility, the fact that PSF coverage tends to be concentrated in the most deprived areas within municipalities, where residents face more important financial constraints and other problems of access to health care services (including those publicly provided), arguably makes it far less likely that a given family would refuse the free PSF services offered to them. Although there seem to be no available statistics on “refusal” rates, a report based on interviews conducted in eight large Brazilian urban centres (Ministerio da Saude, 2005) provides some support to the reasoning above. The report shows that, in all but one of the municipalities included, between 70-93% of the families living in areas covered by the PSF reported receiving at least one completed visit by Family Health Teams each month. Another report shows evidence that the presence of the no-cost PSF services in a given municipality is associated with reduced financial barriers to health care access (Ministerio da Saude, 2006).

"bias stability"[10] suggested by Heckman et al. (1997), which is weaker than unconfoundedness: if the bias generated by the failure of the weak unconfoundedness assumption when comparing individuals living in the treatment region to those in the comparison region can reasonably be assumed to be the same in the periods before and after the programme's implementation, then the estimated ATT for the pre-programme period (i.e., the second term in parentheses in the PSDD estimator (5)) provides an estimate of the bias which can be used to correct the post-programme estimate of the ATT (the first term in parentheses in (5)). The mandatory nature of PSF, coupled with the combined strengths of the matching and difference-in-differences estimators as applied here, make the suggested PSDD-based approach a suitable empirical strategy for evaluating PSF health impacts.

4. Data

My estimations draw on data from two repeated cross-sections of the annual Brazilian Household Survey (*Pesquisa Nacional por Amostra de Domicilios*—PNAD), published by the Brazilian Institute of Geography and Statistics (IBGE). Methodologically, PNAD surveys are three-stage clustered samples where the primary sampling units are the municipalities (a stratified sampling based on the number of residents in the locality), the secondary sampling units are census areas (also a stratified sampling based on the local population) and the tertiary sampling units are the households. The lowest geographic level at which PNAD data are nationally representative is the metropolitan region (MR) (except regarding the rural areas of the North Region until 2004); this is also the lowest level of disaggregation for an individual's place of residence that can be identified in the micro-data.[11]

A number of individual and household socio-economic characteristics are investigated in the fixed modules and sporadic supplements of the PNADs. Questions in the fixed modules are asked in every survey and include household living conditions, demographics, education, labour and income variables. The sporadic supplements on a given theme are usually included at fixed intervals (e.g., five years) and cover issues such as migration, fertility, health, nutrition and child labour. The actual data I use in the empirical work comes from the PNAD health supplements of the 1998 and 2003 waves.[12] These cross-sections cover over 344,000 individuals per year; however, since it is not possible to identify an individual's municipality of residence in the dataset (nor is the dataset representative at that particular level of disaggregation), this study focus on the nine surveyed MRs—Belo Horizonte, Belem, Curitiba, Fortaleza, Porto Alegre, Recife, Rio de Janeiro, Salvador and Sao Paulo—as geographic units for the evaluation of PSF impacts. These are also the main urban areas in the country; together, they represent an overall of 171 municipalities, corresponding in the dataset to more than 127,000 individuals and 34,000 households per wave.

[10] As denominated by Eichler and Lechner (2002).

[11] In the Brazilian context, metropolitan regions correspond to clusters of municipalities usually surrounding—and including—the capitals (or other important municipalities) of a given state. The number of municipalities (and their total population) forming a MR varies greatly; for instance, there are five municipalities in the Belem MR compared to thirty-five municipalities in the Sao Paulo MR. These regions are intended to serve as geographic reference areas only and do not constitute administrative or government levels. The populations of the nine main MRs are almost exclusively urban.

[12] These are the only recent years for which health supplements are available in the PNAD. A health questionnaire was included in the 1981 wave but its comparability relative to the questionnaire used in 1998 and 2003 is severely limited.

The multi-dimensional nature of a programme such as the PSF makes it difficult for the researcher to focus only on very specific or narrow health indicators (e.g., disease-specific ones) when performing an impact evaluation of that intervention on individual health status. PSF has the potential of affecting the health status of individuals at all ages and concerning a number of different health conditions. Fortunately, PNAD health modules include information on some broad health indicators including self-assessed health status, measures of physical mobility and morbidity indicators such as the number of days in bed and inability to perform usual tasks due to illness.

Clearly, using propensity score matching techniques to assess the impact of an intervention on individuals requires a good amount of information on their observable characteristics that can be correlated with the treatment, i.e. exposure to a given PSF coverage level. Arguably, the PNAD datasets meet such requirement. Surveyed persons are asked about household characteristics such as water supply, sewage, waste disposal and electricity; demographics such as gender, age and ethnicity; education characteristics such as literacy, highest degree attained and current school attendance; and other individual variables such as detailed occupational characteristics and income. Thus, data on several potential determinants of individual health status can be used for performing the matching procedures.

Individual data on PSF coverage is not available in the PNADs, nor is it possible to obtain this information from other databases like those published by the Ministry of Health. The only information available from the latter source is the total number of individuals covered by the programme in a given municipality or MR, which is in turn obtained from the information provided by the municipalities on the number of people registered with family health teams. Although the PSF coverage levels observed in a number of Brazilian geographic areas such as municipalities is exactly the kind of data I advocate here as being more suitable for the impact evaluation of health programmes like the PSF, the PNAD individual data (as mentioned above) is representative at the MR level but not at the (lowest) municipality level, and it is not possible to identify an individual's municipality of residence in that dataset. Therefore, for the purpose of this study, it is necessary to use the information on PSF coverage at the MR level in the years 1998 and 2003 to construct the treatment variable. These data are gathered from *Datasus*, the publicly available on-line database of the Brazilian national health system maintained by the Ministry of Health that contains information on the number of individuals registered with family health teams in each MR at the end of the year from 1998 onwards.[13]

The evolution of PSF coverage levels, population health status and socio-economic characteristics in the nine main Brazilian metropolitan regions

There was a notable progress in terms of the number of family health teams formed in the country—and associated PSF coverage levels—between 1994 and 2003. According to the *Datasus*, the number of FHTs in Brazil raised from 328 in 1994 to more than 15,000 at the end of 2003, when around 70% of the country's municipalities had at least one working FHT. The average coverage level in the municipalities that adopted the PSF was around 68% in 2002, yet it masked important differences between states and also between municipalities within states, especially as

[13] This database also contains aggregate information on population health, socio-economic and demographic characteristics. It is available at: http://www.datasus.gov.br. There is no available data on PSF coverage levels in municipalities or MRs prior to 1998.

far as the largest, urban localities are concerned. In particular, during the study period 1998-2003, the paths of growth of PSF coverage levels among the municipalities belonging to the nine main Brazilian MRs have differed significantly after a virtually zero coverage level observed in all of them in 1998 (see Figure 1 and Table 1). For instance, PSF still covered only around 5% of the population living in the Porto Alegre MR in 2003, whilst over 50% of the Belo Horizonte MR population were already covered by the programme in the same year. Moreover, while PSF coverage levels grew steadily over the period 1998-2003 in some MRs like Belo Horizonte and Recife, the observed coverage level experienced a big jump in the case of other MRs such as Belem in 2001 and Curitiba in 2003. Overall, though, the process of PSF implementation in the larger, urban municipalities of the nine main Brazilian MRs seems to have been clearly accelerated by the new mechanism of financial incentives to strategic primary care actions introduced by the federal government in 1998.

INSERT Figure 1

INSERT Table 1

By examining the data on PSF coverage in the nine MRs included in the sample, one can safely consider the year of 1998 as a "before treatment" period for estimation purposes: in that year, the median population coverage by the programme was only 0.6% in the sample (with a maximum observed coverage of 3.5%, corresponding to Belo Horizonte) and it was strictly zero for two MRs (see Table 1). Moreover, with persistently low coverage levels over the entire 1998-2003 period, the second lowest median coverage and the lowest achieved coverage level in the final year (5.3%) among the available MRs, Porto Alegre is the most suitable candidate for serving as the comparison region in the difference-in-differences estimations, representing a situation equivalent to the strict absence of the PSF throughout the study period. It is worth noting that Porto Alegre constitutes one of the most advanced Brazilian regions in socio-economic terms, ranking consistently among the best as far as important indicators of income levels and inequality, poverty, education and population health are concerned (see Table 2).

INSERT Table 2

Due to the multi-dimensional nature of the PSF, it seems reasonable to look at the evolution of indicators that can provide an overall picture of the broad health status of individuals, separately for adults and children. With this aim, three indicators are used in the empirical work as dependent variables to assess the health impacts of alternative PSF coverage levels: (1) self-assessed health, or more specifically, whether the individual reports "very good" or "good" health in a given survey year; (2) whether the individual had been in bed due to illness in the two weeks prior to the survey; and (3) whether the individual had been unable to perform their usual activities due to illness in the two weeks prior to the survey.[14] PSF impacts on these three health outcomes are estimated for two separate sub-samples: adults (for consistency, I use the—arguably broad— IBGE definition of individuals aged 10 years or more, upon which PNAD questions such as those pertaining to employment status are constructed) and children (less than 10 years old).

The precise definitions of the health outcomes and covariates included for performing the propensity score matching procedures are presented in Table 3, while their

[14] These three indicators represent the broadest health measures that can be constructed from the PNADs and are also the health variables for which more information is available in the surveys.

averages across the nine MRs for both sub-samples for the years 1998 and 2003 are shown in Table 4.[15] For adults, in general, the MRs of Belo Horizonte, Rio de Janeiro and Curitiba exhibited the best average health conditions at the beginning of the study judged by the three health indicators investigated, whereas Belem, Fortaleza and Recife presented the worst initial average health conditions. A similar picture emerges as far as the sub-sample of children is concerned, with Porto Alegre (the comparison region) replacing Belo Horizonte among the best performing MRs and Salvador replacing Recife among the worst performing ones. On the one hand, within the sample of nine MRs, the average health condition of adults seems to have deteriorated (or at least not to have improved) in Curitiba, Porto Alegre and Salvador during the period 1998-2003. The same pattern is observed in the case of child health, with the inclusion of Belem among the poorest performers. On the other hand, the MR of Fortaleza clearly stands out as the area where average health conditions improved the most during the period 1998-2003 in the case of all three health indicators, both for adults and children. Interestingly, this is also the MR that exhibited the highest median PSF population coverage level (24%) over the study period; moreover, Porto Alegre, Curitiba and Salvador—the regions where average health conditions did not improve or actually worsened during the study period—were among the four MRs with the lowest observed median PSF coverage levels (in addition to the MR of Rio de Janeiro). The main objective of the empirical application in this paper is to determine whether the aforementioned differences in the evolution of average health status across MRs are due—at least partially—to different levels of PSF population coverage in these MRs over the period 1998-2003, or whether they merely reflect pre-existing differences between the average individuals who reside in each region (e.g., in terms of health endowments).

INSERT Table 3

INSERT Table 4

5. Results

I implement the previously defined ATT estimators (4) and (5) in order to investigate the health impacts of living during the whole study period in one of the eight treatment MRs (Rio de Janeiro, Salvador, Curitiba, Sao Paulo, Belem, Belo Horizonte, Recife and Fortaleza)—that is, being exposed to one of the eight observed (median) PSF coverage levels from 1998 to 2003—compared to living in the comparison MR, Porto Alegre, during the same period (the "no-programme" benchmark).[16] For both the DD and PSDD estimators, one ATT is estimated for each

[15] The usefulness of matching techniques for impact evaluation lies crucially on the amount of information on important covariates that is available to the researcher. Therefore, I started by including in the propensity score estimations the richest set of relevant covariates available in the PNADs; as mentioned above, these included a wealth of household and individual (demographic, education and occupational) characteristics. However, many of these covariates turned out to be far away from statistical significance and were thus withdrawn from the propensity score estimations, on the grounds that including covariates that are only weakly correlated with the treatment variable and/or health outcomes tends to decrease precision (or, more formally, increase the expected mean squared error; see Imbens, 2004). The results for the broadest set of covariates (not shown) are very similar—both in terms of the distribution of propensity scores and ATT point estimates—to those for the restricted set of covariates presented in Table 3.

[16] As it can be seen in Table 1, PSF coverage levels in a given MR varied (in some cases markedly) from one year to the other. In this paper, the empirical results are interpreted vis-à-vis the median PSF coverage level observed over the period 1998-2003, since data on individual health outcomes are only

of the eight relevant pairwise comparisons: being exposed to the PSF coverage level observed in Rio de Janeiro vis-à-vis not exposed to the PSF in Porto Alegre; being exposed to the PSF coverage level observed in Salvador vis-à-vis not exposed to the PSF in Porto Alegre; and so forth. In this context, in order to investigate health effects for treated individuals living in a given MR in 2003, it is important for consistency to consider only residents who have been exposed to the same treatment intensity over the entire study period. Therefore, only individuals who lived in the same MR during 1998-2003 are used in the estimations.[17]

Specification tests

In the case of the PSDD approach, for any given pairwise comparison of PSF exposure levels, there are two non-random individual treatment assignments in my sample of repeated cross-sections: (*i*) MR of residence during 1998-2003 (living in the comparison MR or the corresponding treatment MR) and (*ii*) year of observation (1998 or 2003). Since the distribution of covariates must be the same in the four cells defined by combining these assignments, I follow Blundell et al. (2004) and use a vector of two propensity scores—one probability for each assignment category, conditional on covariates—as matching variables.[18]

Different matching procedures have been initially applied and compared regarding to how well the covariates are balanced across the corresponding sub-samples used to construct each of the three counterfactuals in (5).[19] Simple *t*-tests for differences in means between the relevant treatment and comparison groups before and after matching have been shown to be insufficient for reliably assessing covariate balancing (Imai et al., 2008), so I have also employed other statistical criteria as summary measures to choose between a number of variants of nearest-neighbour, radius and kernel matching procedures. As a general rule, preferred matching methods in a given application are those which exhibit—when contrasting the characteristics of treatment and comparison samples after matching—the lowest median/mean absolute

available for 1998 and 2003 and thus it is not possible to obtain ATT estimates relative to yearly programme coverage levels during the period. The ATT results for residents of a given region could also be interpreted as relative to the particular evolution of coverage levels in the region.

[17] It is not possible in the data to directly determine which individuals stayed in the same MR during 1998-2003, since PNAD migration questions refer either to municipalities or states as geographic units of interest. The safest option is thus to consider only individuals who lived in the same *municipality* during the period of study: these represent approximately 91% of the sample (around 95% of individuals had resided in the same municipality for 3 years or more and only 1% had lived in the municipality for a year or less in 2003; there are only minor differences in these percentages across MRs). Since the bulk of migration in Brazil occurs between municipalities within the same state (this is the case for 61% of the individuals in the sample), it is likely that even more than 91% of the individuals in the data did not actually change their MR of residence during 1998-2003. In view of these numbers, it is arguably the case that migration selectivity driven by differences in PSF coverage levels (with its resulting estimation biases) is unlikely to be a major source of concern during the period of study in my sample of MRs.

[18] For each pairwise comparison of PSF coverage levels, the required propensity scores were estimated in the preferred matching specifications through binary choice models using only the sub-sample of individuals living in one of the two MRs under comparison. See footnote 8.

[19] All the matching alternatives for the PSDD estimations performed in this paper have been implemented imposing joint common support (excluding from the treatment sample those individuals for whom any of the two estimated propensity scores used as matching variables was larger than the corresponding maximum in the relevant comparison sample) and using only the subset of individuals living in the relevant treatment MR in 2003 who were matched for the construction of each and every counterfactual in the ATT estimator (5). Moreover, all specifications were implemented with replacement and allowing for tied nearest neighbours.

standardised bias, lowest pseudo R-squared (for explaining treatment assignment after matching) of the covariates vector used in the estimation of the relevant propensity score and lowest average use of each comparison observation (thus yielding efficiency gains). Additionally, better matching procedures should lead to a relatively small loss in the number of observations in the treatment group so as to maintain the representativeness of the samples for which the ATT is estimated.

According to the above criteria, three methods perform the best in general among all the matching variants attempted, for both the sub-samples of adults and children:[20]

a. *One-to-one nearest neighbour* with replacement and trimming the common support region by excluding 5% of the treatment individuals at which the propensity scores' densities of the comparison observations are the lowest;

b. *Radius matching* with replacement, multiple matches, imposing a calliper excluding 5% of the sample of matched comparison individuals (those 5% of the matched comparison sample who are overall "farther away" in observables relative to their corresponding "treated" individual) and trimming 5% of the treatment sample;

c. *Epanechnikov kernel matching* with bandwidth of 0.01.[21]

The results of the balancing tests for the three preferred matching specifications are shown in Table 5 and Table 6 for adults and children, respectively. For all the alternative methods, the construction of the three required counterfactuals of the PSDD estimator is successful in terms of the achieved balancing of covariates between the relevant treatment and comparison groups. This can be seen by the drastic reductions in the median and mean standardised bias measures after matching and, more intuitively, by the extremely small magnitude of the pseudo R-squared measures, indicating the irrelevance of the set of covariates for explaining an individual's MR of residence—or, equivalently, their exposure to a given PSF

[20] In addition to the two propensity scores (MR of residence and year of observation), the set of matching variables used for balancing treatment and comparison groups in the three preferred specifications also included covariates that tended to remain unbalanced when matching only on the propensity scores, so as to ensure that the distribution of *specific* covariates, as well as their overall distribution, were the same between treatment and comparison groups after matching. Thus, for adults, matching was performed on a vector including the two propensity scores plus three individual covariates (having between 4-7 years of education, living in a household that belonged to the bottom quintile of the distribution of total household income per capita, and living in a household with proper sewage disposal). In contrast, matching directly on covariates (in addition to the two propensity scores) was generally not necessary for achieving either overall or specific covariates balancing in the case of the children sub-samples. For both adults and children, individuals in the relevant treatment and comparison groups were matched based on the Mahalanobis distance between their vectors of matching variables; this measure was also used as the basis for imposing the calliper in radius matching.

[21] The many other matching methods tried here were further variants of one-to-one nearest neighbour, 10-nearest neighbours, radius and kernel matching. For the sake of conciseness, the tests' results for these models are not presented here. The alternatives varied on whether a calliper was imposed for the maximum Mahalanobis distance of comparison individuals in terms of their vector of matching covariates; whether the region of common support was "trimmed" by excluding individuals in the treatment sample; and (in the case of kernel matching) whether Epanechnikov or Gaussian kernel was used and the specific bandwidth chosen. Furthermore, I experimented with estimating the relevant propensity scores using multinomial and ordered choice models (with ordering based on the median PSF coverage level observed in the MR during 1998-2003), but these specifications were normally outperformed in terms of covariates' balancing by those using propensity scores from binary choice models. For more details on the matching alternatives mentioned in this paper, see for instance Cameron and Trivedi (2005).

coverage level—after matching.[22] The common support conditions imposed by each of the matching alternatives typically lead to around 80-90% of the individuals in the original treatment sub-samples being matched to comparison individuals, hence maintaining the representativeness of the samples for each treatment MR used in the PSDD estimations.

INSERT Table 5

INSERT Table 6

Finally, as expected, Table 5 and Table 6 suggest the possibility of performing a sensitivity analysis on the results from alternative matching strategies, by comparing the estimated ATT coming from specifications that favour the use of more information on comparison individuals (radius and kernel matching, which match more comparison individuals in total to a given treated one and exhibit lower average use of each matched comparison, thus increasing efficiency in the ATT estimations) to the estimated results coming from a specification that aims at using only the "best" matches to construct the counterfactuals (nearest neighbour matching, which makes a more intense use of a smaller number of the available comparison individuals and, therefore, sacrifices efficiency in order to diminish potential biases by using "better" matches on average).

Estimates of PSF impacts for adults

The results for the propensity score-difference-in-differences (PSDD) estimator of the ATT described in equation (5) are compared to the estimates from the simple difference-in-differences (DD) estimator (equation (4)). Column (1) of Table 7 presents the results of applying the DD estimator of the ATT to the sub-sample of adults.[23] Overall, higher PSF coverage levels do not seem to lead to better individual health as far as self-assessed health is concerned (Panel A); in fact, most DD point estimates are negative, though statistically significant in only three cases (none of them for the three highest median coverage levels). The only positive point estimate—not significant at conventional levels—corresponds to the ATT of living in Fortaleza, the MR with the highest median coverage level observed during the period: it suggests a small increase of 1.2 percentage point in the probability of an individual reporting good or very good health, compared to what would have happened had the same individual lived in the Porto Alegre MR during the study period.

INSERT Table 7

[22] As demonstrated by Imai et al. (2008), the conclusions of the balancing condition assessment are more trustworthy if a few summary indicators of covariate balancing are examined after matching, instead of relying solely on tests of means as usually done in the applied literature. In this sense, it is reassuring that the results for both adults and children point to major gains in covariate balancing between the relevant treatment and comparison samples after the preferred matching procedures are employed (this adds to the fact that most covariates seem well-balanced after matching according to simple *t*-tests for differences in means, and that exact matching on a few important covariates has been performed in some cases). The preferred balancing assessment procedure suggested by Imai et al. (2008), the separate examination of empirical quantile-quantile plots of the propensity score and/or important matching covariates before and after matching, is impractical for the present empirical exercise due to the large number of matched comparison groups that need to be constructed within a PSDD framework with repeated cross-sections.

[23] The reported standard errors for the DD estimator are robust to heteroskedasticity and clustering at the household level.

As previously discussed, from a theoretical perspective, the estimates coming from the PSDD approach for multiple treatments are preferred to those from the simple DD approach due to the fact that, for a given individual living in the treatment MR, the set of comparison individuals is restricted only to "comparable" individuals in the baseline MR of Porto Alegre, in addition to allowing for bias stability through its DD component. The matching feature of the PSDD estimator seems of particular relevance for the present context where comparison individuals are drawn from a historically well-off Brazilian MR, especially when residents of the treatment group come from relatively deprived MRs such as Belem, Recife and Fortaleza.

In practice, the PSDD results—presented in column (2) of Table 7 for the sub-sample of adults, for the three alternative matching methods—do not differ very much from the DD estimates for the case of self-assessed health.[24] Higher PSF coverage levels still do not seem to consistently result in larger individual probabilities of reporting good or very good health, although positive point estimates become relatively more frequent. The non-significant, positive ATT point estimate of being exposed to the PSF coverage level of Fortaleza MR vis-à-vis the absence of the programme indicates now an increase between 1.4 and 2.9 percentage points in the probability of reporting good or very good health. As expected for this and the remaining outcomes, estimated standard errors are smaller for the less stringent (in the sense of using more comparison individuals in the construction of counterfactuals) radius and kernel matching procedures relative to the more stringent nearest-neighbour method (which uses only the "best" matches).

The results for the other two (narrower) health indicators are, in general, more in accordance with the expected benefits of being exposed to higher PSF coverage levels, yet the corresponding point estimates suggest at best small effects once again. ATT estimates obtained through the DD approach are mostly negative for the probability that an adult had been in bed due to illness in the two weeks previous to the 2003 survey (Panel B), and they tend to become somewhat larger for the two highest PSF coverage levels. According to the estimates in column (1), individuals living in the MRs of Recife and Fortaleza (with median coverage levels above 20%) experienced, respectively, statistically significant probabilities 1.3 and 1.5 percentage points lower on average of having been in bed due to illness, compared to what would have been observed if they had lived in Porto Alegre with no PSF coverage. The preferred PSDD estimates presented in column (2) show a somewhat similar picture, with negative point estimates around one percentage point for most treatment MRs but a larger reduction of between 2.1 and 2.4 percentage points in the case of Fortaleza MR—up to one percentage point larger than the corresponding DD estimate, and coming from both radius matching and the more stringent nearest-neighbour procedure.

[24] Standard errors for the PSDD estimator have been derived analytically under the assumptions of homoskedasticity and independent outcomes across observations belonging to treatment and comparison groups (see, for instance, Eichler and Lechner, 2002). As noted by Imbens (2004), there are no formal results for the variance of the propensity score matching estimators when the propensity score is unknown and needs to be estimated. A common procedure is to estimate standard errors by bootstrapping; however, since the theoretical properties of bootstrap have not yet been established for matching estimators and given the evidence that bootstrapping does not lead to valid confidence intervals for some nearest-neighbour estimators (see Abadie and Imbens, 2006), I have opted for the simpler analytical approximation.

Analogously, there seem to be no increasingly negative effects of higher PSF coverage levels on the probability that individuals had been unable to perform usual tasks due to illness (Panel C), although in general positive PSF coverage levels do seem to result in better health. Apart from an unexpectedly large PSDD estimate for Belem, negative DD and PSDD estimates are normally around one percentage point. Noteworthy exceptions are, once again, the results for the Fortaleza MR: the ATT point estimate from the simple DD approach suggests a statistically significant 2.5 percentage reduction in the corresponding probability relative to the Porto Alegre baseline, whereas the preferred PSDD estimator leads to a larger estimated probability reduction (still statistically significant) between 2.6 and 3.8 percentage points. The PSDD point estimates do not seem to be particularly sensitive to the number of comparison individuals used by the matching procedures, with the less stringent radius and kernel methods reporting the largest and smallest ATT estimates.

Estimates of PSF impacts for children

The ATT estimates for the sub-sample of children are shown in Table 8; in general, for a given health outcome, the estimated treatment effects tend to be larger for children than for adults.

INSERT Table 8

For self-assessed health (Panel A), the estimated programme impacts are usually positive as expected but rarely statistically significant and do not indicate a pattern of increasing health benefits for higher PSF coverage levels, except in the case of the highest observed median coverage corresponding to the Fortaleza MR. For the latter, both the DD and the PSDD approaches agree in producing positive and statistically significant ATT estimates; PSDD results indicate a larger increase in the probability of reporting good or very good health of between 5.1 and 8.4 percentage points for children exposed to the observed PSF coverage in the region, relative to no exposure. In this case, applying stricter criteria to construct the pool of comparison individuals does make a difference for the estimated ATT: a more stringent nearest-neighbour procedure using “best” matches on average leads to a point estimate which is around 3 percentage points larger than those from radius and kernel matching.

Negative point estimates largely predominate for the individual probabilities of having been in bed due to illness (Panel B), yet in the case of the DD estimator they are statistically significant only for the three highest PSF coverage levels, suggesting reductions of between 1.5 and 2 percentage points in the corresponding probability compared to the base case of Porto Alegre residents. The estimates from the PSDD approach are again larger and typically show probability decrements of around 3 percentage points for children living in the MRs of Belo Horizonte, Recife and Fortaleza, compared to similar children living in Porto Alegre. Using nearest-neighbour to include only the most similar comparison individuals in the counterfactuals leads to larger ATT estimates (in absolute value) for Belo Horizonte and Recife. Statistically significant effects are also found for residents of Sao Paulo MR and, counter-intuitively, for the lowest coverage level in the MR of Rio de Janeiro.

Finally, Panel C of Table 8 shows the ATT estimates for the probability of a child having been unable to perform their usual tasks due to illness. The estimated beneficial effects of PSF coverage for this health outcome tend to be even larger than for the previous one, but the patterns in terms of the magnitude and statistical significance of results remain roughly the same. In particular, probability reductions between 2.5 and 4.7 percentage points are found via the simple DD estimator for children living in the three MRs with highest observed coverage levels, compared to what would have happened had those children lived in the Porto Alegre MR. These negative point estimates rise to between 4.4 and 5.3 percentage points if the PSDD estimator is used, with the more stringent nearest-neighbour procedure resulting in a larger ATT estimate for Belo Horizonte MR (by more than 2 percentage points) and smaller point estimates for the MRs of Recife and Fortaleza, compared to radius and kernel methods. Yet again, a clear pattern of increasing health benefits according to PSF coverage levels does not emerge from the empirical results: although negative and statistically significant ATT estimates tend to be found for children living in the three regions with the highest programme coverage levels, relatively large and significant negative point estimates are also found for MRs with low PSF coverage, namely Sao Paulo and Rio de Janeiro.

6. Discussion and conclusions

By estimating average impacts which are independent of a programme's population coverage, most of the empirical literature on the evaluation of public health programmes overlooks the important fact that such interventions create externalities whose magnitude depends crucially on the number of actually treated individuals in a locality, an important component of the observed intensity of treatment. Recent research has convincingly shown that such treatment externalities derived from health programmes exist and can be large enough to severely bias programme impact estimates if not taken into account. However, in general, existing empirical studies are not well-equipped to incorporate these health externalities into their estimated average treatment effects.

To get at these issues, I suggest and apply an empirical approach based on a propensity score-difference-in-differences (PSDD) estimator adapted for the case of multiple treatments. This general approach can be used in the very common non-experimental setting where the interest lies in evaluating the impacts of a health programme for which pre- and post-programme data are available on (*i*) different levels of population coverage across geographic regions (for instance if programme implementation has been phased-in over time and localities), and (*ii*) individuals' health outcomes, socio-economic characteristics and locality of residence. The empirical framework uses the coverage level of the programme in a given area as the treatment variable and has the advantage of incorporating into the estimated average treatment effects any treatment externalities and nonlinearities arising from the level of population coverage of a particular public health intervention. It is also a practical approach in that data on the number of beneficiaries in a given area is often available in real settings, even for developing countries.

The suggested method is applied to evaluate the impacts of exposure to different population coverage levels of the Family Health Programme—PSF (the broadest health programme ever introduced in Brazil) on the health outcomes of adults and children living in eight Brazilian metropolitan regions (MRs), using individual-level data for the years 1998 and 2003. The combined features of matching and difference-in-differences approaches, and the fact that the PSF works as a mandatory programme at the individual level, provide some reassurance that a PSDD estimator is well suited to identify the treatment effects of interest. Statistical tests for the balancing condition indicate that the alternative matching strategies employed succeed in balancing the distribution of covariates across the relevant treatment and comparison sub-samples, without dramatic losses in terms of treatment individuals left unmatched. The different criteria used by each of the matching procedures for constructing the pool of comparison units seem to make a difference in terms of the PSDD estimated treatment effects for children but not for adults. In addition to testing the sensitivity of the preferred PSDD estimates to variations in the way the relevant counterfactuals are constructed, a simple difference-in-differences estimator—which does not impose the common support condition of comparing only the more similar individuals in terms of observables—is also used as a comparator.

Overall, the ATT estimates suggest that positive levels of PSF coverage in a region lead to improvements in individual health outcomes, with relatively small effects for adults but larger estimated impacts for children. The generally positive estimated impacts arising from PSF coverage levels above zero are a plausible result in view of the preliminary evidence which suggests an association between PSF activities and reduced financial barriers to health care access at the municipality level (Goldbaum et al, 2005; Ministerio da Saude, 2006). The largest and statistically significant average treatment effects tend to be found for the residents of metropolitan regions with the three highest median PSF coverage levels during 1998-2003 (Belo Horizonte, Recife and, especially, Fortaleza). Nevertheless, no clear pattern of increasing health benefits according to higher coverage levels emerges from the data; large and statistically significant point estimates are in a few cases also found for individuals living in regions with relatively low coverage levels, such as for children living in Rio de Janeiro.

The generally small PSF impacts found for the three individual health outcomes analysed (and the usual absence of statistically significant effects for the residents of some regions in the sample, e.g., Salvador, Curitiba and Belem) can be explained at least in part by an intrinsic data limitation, namely the low PSF coverage levels observed over the study period for most metropolitan regions. It would have been too optimistic to expect widespread and sizeable health impacts from PSF activities to consistently appear in the estimations for localities where the programme had typically achieved median coverage levels around 5-10% between 1998 and 2003. In only three of the eight MRs in the sample did the median population coverage level reach 15% or more during the period—Belo Horizonte, Recife and Fortaleza. As an intuitive result, the ATT estimates usually point to statistically significant and larger effects for individuals living in the latter MRs (and mainly for Fortaleza, the region with the highest median PSF coverage level observed in the data). These large and significant effects tend to be more frequent for the two most narrowly defined—and, thus, presumably more able to reflect short-term PSF benefits—health outcomes, namely bed episodes and inability to perform usual tasks due to illness (in contrast to self-assessed health).

In light of the above, the often statistically significant and sometimes relatively large estimated impacts for residents of Sao Paulo and Rio de Janeiro MRs are noteworthy exceptions likely due to the existence of other health programmes—run by major municipalities of both MRs—whose population coverage expanded alongside the PSF, thus acting as confounders for the estimated PSF treatment effects in those regions. Due to their political importance and economic power, municipalities such as Sao Paulo and Rio de Janeiro (the capital cities of the homonymous states, which were responsible in each case for more than 54% of the population in the corresponding MRs in 2003) have historically run their health systems in a very autonomous fashion[25], and broad, tailor-made health programmes have been introduced and/or rolled-out in the two capital cities and other large municipalities of these metropolitan regions over the study period. For instance, in Niteroi, one of the largest municipalities of the Rio de Janeiro MR, a health programme akin to the PSF had been in operation since as early as 1992 and continued to be expanded from 1998 to 2003, reaching over 83,000 individuals (nearly 20% of the municipality's population) in the latter year (Hubner and Franco, 2007). In the city of Sao Paulo, the big push to increase the PSF population coverage took place in 2001/2002, yet a wide primary care intervention similar to the PSF and targeting deprived neighbourhoods (*Qualis*, funded by the state and the municipality) was launched in 1996 and has been expanded yearly since then, reaching approximately 690,000 people in 2002—or around 3.7% of the corresponding MR population (Sobrinho and de Sousa, 2002). Since my ATT estimates are based on a difference-in-differences procedure, the fact that increases in PSF coverage in the MRs of Sao Paulo and Rio de Janeiro have probably occurred alongside increases in the population covered by these important co-existing health interventions might have acted as a confounder and—assuming positive health effects and externalities from increased coverage by these concurrent health programmes—have resulted in overestimated treatment effects for such MRs.[26]

An interesting feature for policy purposes to come out from the empirical application are the larger estimated impacts of the programme activities for children vis-à-vis adults, for a given health outcome. This pattern of results provides quantitative evidence suggesting that one of the guidelines of the Brazilian Ministry of Health concerning the operation of the PSF in the municipalities—the focus on child health actions by the Family Health Teams—has been put into practice, having already yielded positive results in the sample of MRs. Corroborative evidence on the prioritisation of child health actions by the PSF has been found at least for ten large Brazilian urban centres (Ministerio da Saude, 2005): in all the studied areas, PSF doctors and nurses indicated child health as their actual priority among the activities specifically encouraged by the Ministry of Health guidelines; likewise, child health and control of hypertension were the two highest priorities identified by community health agents. Moreover, activities such as the monitoring of infant growth and immunisation schemes, and provision of health advice concerning prenatal consultations, breast feeding and oral rehydration therapy by community health agents

[25] For instance, the municipality of Sao Paulo had its own independent health system (*Plano de Atendimento a Saude—PAS*) prior to 2001, only joining the national health system (SUS) in the latter year.

[26] Even though it is not possible, due to insufficient information, to entirely rule out similar confounders for the remaining estimated treatment effects in the paper, health interventions as broad and with important resulting population coverage at the MR level such as those introduced in the municipalities of Rio de Janeiro and Sao Paulo metropolitan regions seem highly unlikely to have been implemented in the municipalities of the poorest MRs, e.g., Recife, Fortaleza, Salvador and Belem. The municipalities of these MRs, located in the relatively deprived Brazilian North and Northeast, have tended to follow more closely the national health system guidance and incentives frameworks regarding local health system organisation, management and health care interventions, mainly due to budgetary, capacity and political constraints which make them more dependent on federal support (see, for instance, Marques and Mendes, 2002 and the references therein). The lack of general capacity by the municipalities of these MRs (including their capital cities) to manage their local health systems in a more "autonomous" fashion and introduce broad, far-reaching alternatives to the federally-sponsored PSF suggests that, even if concurrent health programmes have been implemented and rolled-out from 1998 to 2003 in such regions, they are unlikely to have generated important biases for the respective estimated treatment effects.

were rated good/excellent by at least three quarters of the interviewed families. Improved maternal management of diarrhoea and respiratory infections, in addition to higher rates of immunisation and breast feeding, were also found in those areas.

Although the simple difference-in-differences (DD) and the preferred PSDD estimates broadly agree as far as the sign and statistical significance of point estimates are concerned, the latter estimator tends to yield somewhat larger treatment effects than the former. One likely reason for this pattern lies in the average characteristics of Porto Alegre MR residents (the pool of comparison units) relative to those living in the eight treatment MRs, and the different ways each of two estimators take such differences into account. As shown in the data description section, the Porto Alegre MR ranks consistently among the two or three most advanced regions in the sample in terms of important socio-economic indicators, for both 1998 and 2003. The average socio-economic contrasts are small for MRs like Sao Paulo and Curitiba but can be substantial when comparisons of Porto Alegre residents are made against individuals living in the most disadvantaged treatment MRs in the sample, such as Fortaleza, Recife and Salvador. The fact that residents of Porto Alegre exhibit, on average, considerably better socio-economic conditions than those living in, say, Fortaleza, means that comparisons of the evolution of health outcomes within the simple DD approach often involve very dissimilar individuals in terms of observables.[27] DD treatment effects are linearly extrapolated outside the common support region, i.e. the estimates of individual treatment effects for the Fortaleza residents with only a few or no "similar" comparison individuals in Porto Alegre are basically linear extrapolations of the results found for the rest of the sample (the "more comparable" units). If PSF health effects are non-linear, this could introduce some bias in the estimate of *average* treatment effects: for instance, if the more socially disadvantaged individuals living in Fortaleza—for whom there are relatively few "comparable" counterparts in Porto Alegre—tend to benefit more from the preventive actions encouraged by PSF professionals (even if only through the externality component of interacting with actually covered people in the region), the simple DD estimator could lead to an underestimation of the average treatment effect of being exposed to the PSF in Fortaleza.

The PSDD estimator, in contrast, has the advantage of forcing the treatment effect estimate for a given treated resident to come from a comparison only with "similar" Porto Alegre MR residents (with variable degrees of similarity in observables imposed by the researcher through the chosen matching method), explicitly taking the aforementioned non-linearities into account and thus providing more reliable point estimates of average treatment effects. Therefore, in my data, the often larger magnitude of PSDD point estimates relative to the simple DD results seems to be a symptom of the (sometimes substantial) differences between average residents of the comparison and treatment MRs, with the simple DD estimator probably providing underestimated ATT estimates for residents of the most socially disadvantaged MRs.[28]

---

[27] This adds to the already mentioned prioritisation of the poorest and unhealthiest areas within the municipalities when it comes to the allocation of Family Health Teams, so the individuals *actually visited* by PSF teams tend to be the worst-off within a given MR.

[28] It is worth emphasising that the PSDD estimator accounts for non-linearities in treatment effects as well as the lower average health endowments in these socially disadvantaged treatment MRs compared to Porto Alegre (through its difference-in-differences component).

In sum, the generally positive health impact estimates for residents of regions with positive PSF coverage levels suggest that primary care interventions based on multi-professional teams which perform and encourage basic preventive health activities have the potential of improving health status, through direct provision of care and its associated externalities. This basic result—found specifically for the Brazilian case and which might be relevant for the context of other less-developed countries—adds to preliminary evidence elsewhere pointing towards increased access to health services in areas covered by the PSF. However, as with any other approach based on non-experimental methods, the reliability of the results from the evaluation strategy applied in this paper depends crucially on the quality of the available data. From the discussion so far, clear limitations of the data used in this paper for evaluating PSF health impacts are, for instance, the relatively low population coverage levels by the programme observed in Brazilian metropolitan regions and the reduced number of different coverage levels included in the analysis (eight). It would have been desirable for empirical purposes to have a variety of areas with population coverage levels ranging from close to zero to close to 100%, so as to estimate average impacts of marginally increasing PSF coverage levels and identify increasing/diminishing returns to expanded coverage.[29]

This paper provides one of the few empirical applications of recent extensions of the propensity-score matching estimator for the case of multiple treatments. Despite the important data limitations faced by the empirical application, the potential advantages of the PSDD-based approach suggested here clearly make it a useful tool for policy purposes. The health programme evaluation strategy used in this paper can be applied in the very frequent situation (especially in a developing country context) where information on the actual impacts of a health intervention is needed to guide resource allocation and roll-out strategies—but only routine, non-experimental data are available to the researcher. Furthermore, the estimates of programme impacts incorporate the (potentially substantial) treatment non-linearities and externalities arising from different levels of population coverage. Computing the average impact of a health intervention on the general population resident in the area—and not only its effects on actually "treated" individuals—will contribute to provide a more comprehensive account of the health benefits generated by the intervention and produce a result which is probably of more relevance for policy-makers.

References

Abadie, A., and G. Imbens (2006). "On the failure of the bootstrap for matching estimators". NBER Technical Working Paper 325.

Angeles, G., Guilkey, D. & Mroz, T. (2005) "The Determinants of Fertility in Rural Peru: Program Effects in the Early Years of the National Family Planning Program", Journal of Population Economics 18(2): 367-389.

[29] When the interest lies in evaluating the impacts of a health programme for which data are available on several different levels of population coverage across geographic regions, a more sensible evaluation strategy can be the utilisation of propensity score methods adapted to the continuous treatment case, in contrast to the multiple treatments setting suggested here. It has been only recently that a methodological contribution for the continuous treatment context has been advanced in the literature by Hirano and Imbens (2004), who discuss and illustrate the application of propensity scores in a regression context for evaluating the impact of continuous treatments. For an empirically oriented review of these methods, see Moreno-Serra (2007).

Armecin, G., J. Behrman, P. Duazo, S. Ghuman, S. Gultiano, E. King and N. Lee (2006). “Early Childhood Development through an Integrated Program: Evidence from the Philippines”, World Bank Policy Research Working Paper 3922, May.

Attanasio, O., & Vera-Hernández, M. (2004) “Medium and Long Run Effects of Nutrition and Child Care: Evaluation of a Community Nursery Programme in Rural Colombia”, EDePo-IFS Working Paper 04-06.

Blundell, R., Costa-Dias, M. (2000) “Evaluation Methods for Non-Experimental Data”, Fiscal Studies 21(4): 427-468.

Blundell, R., M. Costa-Dias, C. Meghir and J. Van Reenen (2004) “Evaluating the Employment Impact of a Mandatory Job Search Program”, Journal of the European Economic Association 2(4): 569-606.

Cameron, A. and P. Trivedi (2005). Microeconometrics: Methods and Applications. Cambridge, UK: Cambridge University Press.

Eichler, M. and M. Lechner (2002) “An evaluation of public employment programmes in the East German State of Sachsen-Anhalt”, Labour Economics 9: 143-186.

Frankenberg, E., Suriastini, W., Thomas, D. (2005) “Can Expanding Access to Basic Healthcare Improve Children’s Health Status? Lessons from Indonesia’s ‘Midwife in the Village’ Programme”, Population Studies 59(1): 5-19.

Goldbaum, M., RJ Gianini, HM Novaes, et al (2005). “Health services utilization in areas covered by the family health program (Qualis) in Sao Paulo City, Brazil”. Revista de Saúde Pública 39: 90–99.

Heckman, J., Ichimura, H. and Todd, P. (1997) “Matching as an econometric evaluation estimator: evidence from evaluating a job training programme”, Review of Economic Studies 64: 605-654.

Heckman, J., Smith, J. (1995) “Assessing the Case for Social Experiments”, Journal of Economic Perspectives 9(2): 85-110.

Hirano, K. and G. Imbens (2004). “The propensity score with continuous treatments”. In: A. Gelman and X.L. Meng (Eds.), Missing Data and Bayesian Methods in Practice. New York: Wiley.

Hubner, L. and T. Franco (2007). “O Programa Médico de Família de Niterói como estratégia de implementação de um modelo de atenção que contemple os princípios e diretrizes do SUS”. PHYSIS: Revista de Saúde Coletiva 17(1): 173-191.

Imai, K., G. King and E. Stuart (2008) “Misunderstandings between experimentalists and observationalists about causal inference”. Journal of the Royal Statistical Society Series A, 171:2, 481-502.

Imbens, G. (2004). “Nonparametric estimation of average treatment effects under exogeneity: A review”. Review of Economics and Statistics 86(1): 4-29.

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

Jalan, J. and M. Ravallion (2003) “Does piped water reduce diarrhea for children in rural India?”, Journal of Econometrics 112: 153-173.

Macinko, J. Guanais, F. & F. Souza. (2006). An Evaluation of the Impact of the Family Health Program on Infant Mortality in Brazil, 1990-2002. Journal of Epidemiology and Community Health 60:13-19.

Marques, R. and A. Mendes (2002). "A dimensão do financiamento da atenção básica e do PSF no contexto da saúde—SUS". In: de Sousa, M. (Ed.), Os Sinais Vermelhos do PSF. São Paulo: Hucitec, pp. 71-101.

Miguel, E., Kremer, M. (2004) "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities". Econometrica 72(1): 159–217.

Ministério da Saúde (2006). Saúde da família no Brasil: Uma análise de indicadores selecionados: 1998-2004. Secretaria de Atenção à Saúde, Departamento de Atenção Básica. Brasília, D.F.: Ministério da Saúde.

Ministério da Saúde (2005). Saúde da Família: Avaliação da implementação em dez grandes centros urbanos: síntese dos principais resultados. Ministério da Saúde e Fundação Oswaldo Cruz. Brasília, D.F.: Ministério da Saúde.

Ministério da Saúde (2004). Avaliação normativa do Programa Saúde da Família no Brasil: monitoramento da implantação e funcionamento das equipes de saúde da família: 2001-2002. Secretaria de Atenção à Saúde, Departamento de Atenção Básica. Brasília, D.F.: Ministério da Saúde.

Ministério da Saúde (2001). Programa Saúde da Família. Secretaria Executiva. Brasília, D.F.: Ministério da Saúde.

Moreno-Serra, R. (2007). "Matching estimators of average treatment effects: a review applied to the evaluation of health care programmes", HEDG Working Paper 07/02, February.

Moreno-Serra, R. (2005) "Uma avaliação do impacto do Programa Saúde da Família sobre a saúde infantil no estado de São Paulo". In S. Piola and E. Jorge (Eds.), Economia da Saúde: 1o Prêmio Nacional—2004. Coletânea Premiada. Brasília: IPEA-DFID, pp. 79-112.

Sobrinho, E. and M. de Sousa (2002). "O PSF nos grandes municípios: São Paulo, um investimento à parte!". In: M. de Sousa (Ed.), Os Sinais Vermelhos do PSF. São Paulo: Hucitec, pp. 47-68.

Smith, J. & Todd, P. (2005) "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?", Journal of Econometrics 125: 305-353.

Table 1: PSF coverage levels in 9 metropolitan regions, 1998-2003 (% population)

Metropolitan Region	1998	1999	2000	2001	2002	2003	Median
Belem	0.0	0.7	2.4	20.4	22.4	22.7	11.4
Belo Horizonte	3.5	5.6	13.3	19.3	42.0	51.1	16.3
Curitiba	2.9	3.3	3.7	14.3	6.1	28.5	4.9
Fortaleza	0.6	10.4	26.0	21.9	35.8	35.9	24.0
Porto Alegre	2.7	3.1	3.4	4.5	5.2	5.3	4.0
Recife	0.3	8.9	17.8	27.9	34.0	36.9	22.8
Rio de Janeiro	0.3	0.4	1.6	4.1	7.7	8.6	2.9
Salvador	0.0	1.9	3.0	5.1	6.8	8.6	4.0
Sao Paulo	0.9	1.7	4.9	6.8	10.7	14.0	5.9

Note: Calculated based on Datasus official figures for the number of individuals registered with Family Health Teams in December of the corresponding year.

Table 2: Aggregate socio-economic indicators, 9 metropolitan regions (1998-2003)

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
		GDP per capita (R$)	Income ratio	Poor (%)	Unemployment rate	Illiteracy rate	Eight or more years of education (%)	Households with piped water (%)	Households with waste disposal (%)	Infant mortality rate	Low birthweight (%)	Death rate Cerebrovascular diseases	Death rate Respiratory diseases	Hospital beds per 1,000	Health professionals per 1,000
Year 1998															
	Belem	4,737	25.0	31.4	10.3	4.9	52.0	80.8	93.5	26.7	7.8	55.5	58.1	2.5	11.2
	Belo Horizonte	10,475	22.8	23.4	12.7	6.3	42.9	96.3	89.8	22.9	10.2	51.2	73.5	3.0	16.9
	Curitiba	11,216	20.8	15.3	11.2	4.6	49.9	91.0	96.6	20.6	8.9	53.4	73.0	3.5	13.6
	Fortaleza	5,774	23.3	41.4	11.0	15.2	40.2	80.5	90.4	26.8	7.5	42.0	45.5	2.8	12.8
	Porto Alegre	12,329	20.9	14.6	11.1	4.5	47.8	90.3	97.2	16.2	8.8	71.1	102.0	3.1	15.0
	Recife	7,120	28.0	40.6	14.7	12.0	43.2	89.8	86.0	23.3	8.2	71.0	75.2	3.7	16.3
	Rio de Janeiro	10,751	20.2	13.9	11.1	4.5	52.2	89.0	92.6	20.9	9.1	81.6	94.6	3.6	17.1
	Salvador	11,320	27.7	31.8	17.2	7.7	49.8	94.6	92.5	24.4	9.6	52.2	55.1	3.0	18.2
	Sao Paulo	13,773	20.9	12.1	14.9	5.3	52.3	98.1	98.4	20.0	8.8	54.4	67.3	2.3	14.0
Year 2003															
	Belem	4,614	18.4	43.4	11.9	5.6	55.5	65.2	95.7	21.9	9.0	47.9	59.3	2.5	12.9
	Belo Horizonte	10,677	19.9	26.4	11.7	5.9	55.5	98.4	96.4	15.9	11.3	46.9	53.5	2.7	17.5
	Curitiba	10,849	17.9	18.4	9.4	3.6	59.8	93.7	97.2	15.4	9.9	51.7	55.7	3.1	15.6
	Fortaleza	5,510	21.7	47.8	13.6	12.0	50.0	85.6	91.8	23.3	7.9	40.0	47.3	2.5	12.3
	Porto Alegre	12,842	19.6	21.1	10.0	3.8	56.3	88.2	98.3	14.0	9.9	66.2	76.9	2.7	17.0
	Recife	7,645	27.3	46.2	17.5	11.4	50.3	89.1	93.9	17.3	8.9	63.3	69.1	3.5	18.0
	Rio de Janeiro	9,536	20.6	19.8	13.6	3.6	59.1	91.1	98.2	17.0	9.5	71.2	85.1	3.3	17.2
	Salvador	11,786	29.5	43.3	19.8	7.1	57.3	97.3	96.8	24.1	10.4	45.2	53.9	2.7	18.1
	Sao Paulo	12,961	23.4	20.2	14.6	4.6	60.8	97.9	99.2	15.2	9.6	48.1	69.0	2.1	14.2

Notes: (1) Real GDP per capita in Brazilian reais (R$), 2000 prices; (2) Total income of individuals belonging to the top per capita income quintile divided by the total income of individuals in the bottom per capita income quintile; (3) Proportion of the resident population with per capita household income up to 1/2 minimum wage; (4) Proportion of the labour force without a job (individuals aged at least 10 years old who are actively looking for employment); (5) Proportion of individuals aged 15 years or more who cannot read and write; (6) Proportion of the resident population aged 15 years or more with eight or more years of schooling; (7) Proportion of households connected to piped water; (8) Proportion of households served by waste collection and disposal systems; (9) Number of deaths among children aged 1 year or less (per 1,000 births); (10) Proportion of low-birthweight newborns (less than 2,5kg); (11) Crude death rate for cerebrovascular diseases (number of deaths per 100,000 residents); (12) Crude death rate for respiratory diseases (number of deaths per 100,000 residents); (13) Total number of hospital beds per 1,000 residents; (14) Total number of health professionals per 1,000 residents.

Table 3: Variable definitions

Variable	Sub-samples	Definition
Good/Very good self-assessed health	Adults/Children	Dummy variable taking on the value of 1 if the individual reported either "very good" or "good" health in the date of the survey; 0 otherwise. Constructed based on the individual answers to the question "Overall, you consider your own health satus to be: very good/good/fair/poor/very poor". The question was answered by the parents/legal guardians in the case of children.
Bed due to illness	Adults/Children	Dummy variable taking on the value of 1 if the individual had been in bed due to illness in the two weeks prior to the survey; 0 otherwise. Constructed based on the individual answers to the question "In the last two weeks, have you been in bed due to health reasons?". The question was answered by the parents/legal guardians in the case of children.
Unable to perform usual tasks due to illness	Adults/Children	Dummy variable taking on the value of 1 if the individual had been unable to perform their usual activities due to illness in the two weeks prior to the survey; 0 otherwise. Constructed based on the individual answers to the question "In the last two weeks, have you been unable to perform any of your usual activities (work, attend school, play etc.) due to health reasons?". The question was answered by the parents/legal guardians in the case of children.
Age	Adults/Children	Age in years.
Black	Adults/Children	Dummy variable taking on the value of 1 if the individual reports their colour/race to be Black (*preta*); 0 otherwise. White is the base category.
Mixed	Adults/Children	Dummy variable taking on the value of 1 if the individual reports their colour/race to be Mixed/Brown (*parda*); 0 otherwise. White is the base category.
Other ethnicity	Adults/Children	Dummy variable taking on the value of 1 if the individual reports their colour/race to be either Amerindian (*indigena*) or Asian (*amarela*); 0 otherwise. White is the base category.
Male	Adults/Children	Dummy variable taking on the value of 1 if the individual is male; 0 otherwise.
Illiterate	Only adults	Dummy variable taking on the value of 1 if the individual can read and write; 0 otherwise.
Unemployed	Only adults	Dummy variable taking on the value of 1 if the individual is unemployed; 0 otherwise. The definition of unemployment considered here is wider than the usual definition of the Brazilian Institute of Geography and Statistics (IBGE) and corresponds to a "relaxed definition" often advocated for the case of developing countries in the labour literature: the criterion of "seeking work" is relaxed to take into account the important prevalence of elements such as "hidden unemployment" (individuals in informal work etc.) and "discouraged workers". See, for instance, ILO's *Key Indicators of the Labour Market* publications.
0-3 years of education	Only adults	Dummy variable taking on the value of 1 if the individual has less than 4 years of formal education; 0 otherwise. Eleven or more years of formal education is the base category.
4-7 years of education	Only adults	Dummy variable taking on the value of 1 if the individual has between 4 and 7 years of formal education (inclusive); 0 otherwise. Eleven or more years of formal education is the base category.

Variable	Sub-samples	Definition
8-10 years of education	Only adults	Dummy variable taking on the value of 1 if the individual has between 8 and 10 years of formal education (inclusive); 0 otherwise. Eleven or more years of formal education is the base category.
Kindergarten	Only children	Dummy variable taking on the value of 1 if the child is currently attending kindergarten or nursery; 0 otherwise. Not attending any kind of school is the base category.
Basic school	Only children	Dummy variable taking on the value of 1 if the child is currently attending school and enrolled in the first, second, third or fourth year of basic formal education (*ensino fundamental/primeiro grau*); 0 otherwise. Not attending any kind of school is the base category.
First income quintile	Adults/Children	Dummy variable taking on the value of 1 if the individual lives in a household that belongs to the bottom quintile of the distribution of total household income per capita in the entire sample of metropolitan regions; 0 otherwise. Living in a household in the top quintile of the distribution is the base category.
Second income quintile	Adults/Children	Dummy variable taking on the value of 1 if the individual lives in a household that belongs to the second quintile of the distribution of total household income per capita in the entire sample of metropolitan regions; 0 otherwise. Living in a household in the top quintile of the distribution is the base category.
Third income quintile	Adults/Children	Dummy variable taking on the value of 1 if the individual lives in a household that belongs to the third quintile of the distribution of total household income per capita in the entire sample of metropolitan regions; 0 otherwise. Living in a household in the top quintile of the distribution is the base category.
Fourth income quintile	Adults/Children	Dummy variable taking on the value of 1 if the individual lives in a household that belongs to the fourth quintile of the distribution of total household income per capita in the entire sample of metropolitan regions; 0 otherwise. Living in a household in the top quintile of the distribution is the base category.
Piped water	Adults/Children	Dummy variable taking on the value of 1 if the individual lives in a household connected to piped water; 0 otherwise.
Sewage	Adults/Children	Dummy variable taking on the value of 1 if the individual lives in a household with proper sewage disposal; 0 otherwise.
Residents	Adults/Children	Number of residents in the individual's household.

Table 4: Sample averages by metropolitan region: adults and children (1998 & 2003)

Panel A: Adults

Year 1998

Variable	Belem	Belo Horizonte	Curitiba	Fortaleza	Porto Alegre	Recife	Rio de Janeiro	Salvador	Sao Paulo
Good/Very good self-assessed health	0.720	0.811	0.794	0.748	0.791	0.711	0.810	0.764	0.790
Bed due to illness	0.051	0.034	0.034	0.045	0.038	0.041	0.033	0.037	0.040
Unable to perform usual tasks due to illness	0.089	0.060	0.058	0.076	0.068	0.068	0.049	0.064	0.065
Age	33.5	33.9	34.3	33.3	36.1	33.7	37.5	32.4	35.2
Black	0.036	0.090	0.028	0.017	0.081	0.072	0.123	0.204	0.059
Mixed	0.666	0.414	0.140	0.653	0.089	0.493	0.260	0.595	0.246
Other ethnicity	0.003	0.006	0.009	0.001	0.003	0.001	0.001	0.005	0.028
Male	0.463	0.475	0.487	0.465	0.471	0.461	0.463	0.465	0.472
Illiterate	0.047	0.059	0.039	0.140	0.043	0.106	0.041	0.072	0.046
0-3 years of education	0.241	0.216	0.196	0.336	0.194	0.279	0.210	0.238	0.187
4-7 years of education	0.304	0.405	0.354	0.319	0.393	0.341	0.317	0.329	0.348
8-10 years of education	0.192	0.165	0.194	0.151	0.171	0.155	0.191	0.170	0.204
Unemployed	0.512	0.471	0.451	0.496	0.449	0.549	0.518	0.517	0.499
First income quintile	0.286	0.200	0.129	0.388	0.127	0.386	0.121	0.298	0.110
Second income quintile	0.220	0.234	0.169	0.241	0.182	0.232	0.198	0.234	0.143
Third income quintile	0.171	0.215	0.220	0.144	0.223	0.148	0.237	0.184	0.210
Fourth income quintile	0.176	0.184	0.235	0.119	0.237	0.118	0.228	0.141	0.270
Piped water	0.880	0.978	0.984	0.826	0.982	0.915	0.968	0.933	0.993
Sewage	0.396	0.845	0.687	0.334	0.856	0.430	0.859	0.679	0.914
Residents	5.5	4.6	4.2	5.1	4.0	4.9	3.9	4.9	4.4
Observations	4,609	11,332	6,323	9,304	12,153	10,528	14,349	9,794	14,664

Panel A: Adults (contd.)

Year 2003

Variable	Belem	Belo Horizonte	Curitiba	Fortaleza	Porto Alegre	Recife	Rio de Janeiro	Salvador	Sao Paulo
Good/Very good self-assessed health	0.700	0.818	0.788	0.780	0.812	0.720	0.823	0.735	0.798
Bed due to illness	0.050	0.033	0.040	0.035	0.044	0.036	0.030	0.037	0.037
Unable to perform usual tasks due to illness	0.087	0.056	0.061	0.056	0.074	0.066	0.041	0.062	0.059
Age	33.8	35.9	35.5	33.8	37.4	35.2	39.1	33.5	36.2
Black	0.056	0.103	0.031	0.032	0.080	0.070	0.110	0.269	0.065
Mixed	0.682	0.419	0.151	0.621	0.069	0.554	0.295	0.538	0.263
Other ethnicity	0.005	0.003	0.009	0.006	0.003	0.010	0.002	0.008	0.021
Male	0.473	0.471	0.484	0.474	0.471	0.462	0.455	0.474	0.470
Illiterate	0.053	0.055	0.034	0.108	0.036	0.108	0.033	0.066	0.040
0-3 years of education	0.202	0.172	0.171	0.254	0.154	0.238	0.171	0.210	0.157
4-7 years of education	0.303	0.334	0.298	0.312	0.349	0.318	0.291	0.285	0.293
8-10 years of education	0.211	0.176	0.199	0.172	0.184	0.156	0.197	0.190	0.192
Unemployed	0.478	0.458	0.440	0.509	0.437	0.560	0.521	0.499	0.482
First income quintile	0.193	0.101	0.079	0.259	0.066	0.267	0.076	0.242	0.092
Second income quintile	0.239	0.159	0.117	0.228	0.108	0.212	0.121	0.215	0.104
Third income quintile	0.237	0.245	0.202	0.215	0.180	0.218	0.204	0.216	0.182
Fourth income quintile	0.186	0.246	0.293	0.156	0.296	0.161	0.274	0.159	0.274
Piped water	0.856	0.995	0.983	0.872	0.986	0.922	0.982	0.961	0.995
Sewage	0.388	0.859	0.781	0.542	0.842	0.424	0.874	0.805	0.882
Residents	5.0	4.1	4.1	4.6	3.8	4.4	3.7	4.5	4.1
Observations	8,246	9,058	5,660	10,839	14,071	11,928	13,659	11,629	16,615

Panel B: Children (less than 10 years old)

Year 1998

Covariate	Belem	Belo Horizonte	Curitiba	Fortaleza	Porto Alegre	Recife	Rio de Janeiro	Salvador	Sao Paulo
Good/Very good self-assessed health	0.802	0.924	0.910	0.894	0.924	0.866	0.943	0.873	0.935
Bed due to illness	0.061	0.037	0.024	0.052	0.038	0.041	0.035	0.047	0.044
Unable to perform usual tasks due to illness	0.115	0.072	0.056	0.091	0.061	0.080	0.045	0.080	0.076
Age	4.6	4.6	4.4	4.7	4.5	4.5	4.6	4.6	4.6
Black	0.018	0.073	0.019	0.009	0.077	0.053	0.126	0.142	0.038
Mixed	0.646	0.445	0.174	0.666	0.125	0.520	0.303	0.695	0.283
Other ethnicity	0.002	0.006	0.001	0.002	0.002	0.001	0.000	0.003	0.014
Male	0.495	0.491	0.502	0.511	0.517	0.511	0.504	0.518	0.518
Kindergarten	0.230	0.227	0.159	0.308	0.161	0.312	0.276	0.286	0.227
Basic school	0.276	0.270	0.278	0.247	0.281	0.257	0.260	0.256	0.270
First income quintile	0.466	0.372	0.252	0.587	0.276	0.566	0.269	0.482	0.208
Second income quintile	0.215	0.255	0.252	0.209	0.244	0.202	0.241	0.239	0.218
Third income quintile	0.145	0.173	0.215	0.099	0.221	0.099	0.224	0.128	0.219
Fourth income quintile	0.104	0.112	0.163	0.060	0.162	0.077	0.159	0.083	0.206
Piped water	0.797	0.969	0.969	0.775	0.957	0.843	0.952	0.890	0.990
Sewage	0.331	0.784	0.603	0.277	0.809	0.347	0.812	0.613	0.882
Residents	6.4	5.0	4.8	5.7	4.9	5.4	4.7	5.4	5.0
Observations	938	2,468	1,475	2,461	2,553	2,396	2,538	2,000	2,675

Panel B: Children (less than 10 years old) (contd.)

Year 2003

Covariate	Belem	Belo Horizonte	Curitiba	Fortaleza	Porto Alegre	Recife	Rio de Janeiro	Salvador	Sao Paulo
Good/Very good self-assessed health	0.814	0.944	0.917	0.929	0.934	0.874	0.957	0.833	0.918
Bed due to illness	0.080	0.031	0.047	0.041	0.048	0.038	0.030	0.054	0.041
Unable to perform usual tasks due to illness	0.153	0.075	0.086	0.076	0.090	0.085	0.042	0.088	0.069
Age	4.5	4.9	4.8	4.8	4.7	4.7	4.8	4.8	4.7
Black	0.037	0.079	0.024	0.024	0.083	0.038	0.111	0.235	0.046
Mixed	0.675	0.459	0.166	0.600	0.105	0.566	0.381	0.592	0.301
Other ethnicity	0.003	0.001	0.005	0.004	0.002	0.007	0.001	0.006	0.007
Male	0.523	0.534	0.539	0.501	0.515	0.509	0.503	0.518	0.509
Kindergarten	0.278	0.290	0.233	0.344	0.198	0.330	0.307	0.311	0.270
Basic school	0.255	0.304	0.337	0.284	0.284	0.297	0.283	0.306	0.297
First income quintile	0.320	0.221	0.173	0.449	0.187	0.466	0.210	0.418	0.212
Second income quintile	0.284	0.250	0.204	0.238	0.208	0.223	0.229	0.245	0.174
Third income quintile	0.214	0.235	0.243	0.169	0.230	0.154	0.218	0.164	0.213
Fourth income quintile	0.110	0.169	0.244	0.086	0.214	0.101	0.198	0.098	0.221
Piped water	0.776	0.992	0.979	0.823	0.969	0.887	0.969	0.930	0.990
Sewage	0.314	0.813	0.736	0.491	0.780	0.361	0.830	0.759	0.814
Residents	5.7	4.6	4.9	5.0	4.8	5.0	4.5	5.2	4.8
Observations	1,906	1,689	1,144	2,467	2,684	2,345	2,168	2,398	3,098

Table 5: Tests for balancing of covariates for the three preferred matching specifications, before and after matching: adults

Counterfactual (1): Comparable adults living in Porto Alegre in 2003

Matching method	Absolute standardised bias				Pseudo R-squared		Treated group			Comparison group		
	Median		Mean				Observations			Observations		Average use
	Before	After	Before	After	Before	After	Before	After	Lost	Total	Used	
Nearest-neighbour												
Rio de Janeiro	4.91	0.39	9.05	0.55	0.089	0.000	12,748	11,742	1,006	13,563	4,858	2.4
Salvador	13.93	0.28	25.40	1.31	0.396	0.003	10,916	9,883	1,033	13,563	2,868	3.4
Curitiba	3.44	0.08	7.10	0.46	0.038	0.001	5,477	5,055	422	13,563	3,233	1.6
Sao Paulo	4.89	0.13	9.39	0.51	0.075	0.001	15,630	14,435	1,195	13,563	5,391	2.7
Belem	17.12	0.41	30.82	1.70	0.451	0.006	7,258	6,145	1,113	13,563	2,039	3.0
Belo Horizonte	5.64	0.05	11.69	0.46	0.145	0.001	8,686	7,878	808	13,563	3,661	2.2
Recife	23.20	0.43	29.27	1.58	0.372	0.003	10,749	9,569	1,180	13,563	3,133	3.1
Fortaleza	23.63	0.51	31.15	1.53	0.403	0.003	10,116	9,064	1,052	13,563	2,810	3.2
Radius												
Rio de Janeiro	5.35	0.48	9.42	0.91	0.083	0.001	12,748	10,665	2,083	13,563	12,126	0.9
Salvador	11.91	1.29	23.63	2.62	0.393	0.003	10,916	8,936	1,980	13,563	12,628	0.7
Curitiba	5.05	1.41	7.14	1.62	0.039	0.002	5,477	4,769	708	13,563	11,689	0.4
Sao Paulo	6.37	1.46	9.69	1.42	0.074	0.002	15,630	13,399	2,231	13,563	12,377	1.1
Belem	19.00	2.49	30.23	2.56	0.443	0.003	7,258	5,548	1,710	13,563	13,345	0.4
Belo Horizonte	6.00	1.08	11.88	1.29	0.141	0.002	8,686	7,261	1,425	13,563	12,331	0.6
Recife	21.38	1.72	28.31	3.14	0.360	0.006	10,749	8,561	2,188	13,563	11,774	0.7
Fortaleza	21.50	1.90	30.06	2.92	0.395	0.005	10,116	8,036	2,080	13,563	11,299	0.7
Kernel												
Rio de Janeiro	6.35	0.56	9.84	1.24	0.090	0.001	12,748	11,688	1,060	13,563	12,898	0.9
Salvador	11.56	1.51	20.93	2.30	0.358	0.003	10,916	8,396	2,520	13,563	12,447	0.7
Curitiba	5.33	1.20	7.56	1.32	0.030	0.001	5,477	4,357	1,120	13,563	10,914	0.4
Sao Paulo	4.81	1.11	8.60	1.34	0.058	0.001	15,630	13,299	2,331	13,563	12,482	1.1
Belem	9.10	1.40	22.64	2.22	0.307	0.003	7,258	3,122	4,136	13,563	11,290	0.3
Belo Horizonte	6.47	0.84	11.78	1.25	0.143	0.001	8,686	7,678	1,008	13,563	12,568	0.6
Recife	19.63	1.61	25.32	2.50	0.264	0.005	10,749	7,034	3,715	13,563	12,200	0.6
Fortaleza	14.78	3.57	23.40	3.05	0.260	0.004	10,116	5,932	4,184	13,563	11,815	0.5

Counterfactual (2): Comparable adults living in the same MR in 1998

Matching method	Absolute standardised bias				Pseudo R-squared		Treated group			Comparison group		
	Median		Mean				Observations			Observations		Average use
	Before	After	Before	After	Before	After	Before	After	Lost	Total	Used	
Nearest-neighbour												
Rio de Janeiro	5.32	0.13	7.37	0.36	0.028	0.000	12,748	11,742	1,006	13,494	5,570	2.1
Salvador	6.44	0.10	8.57	0.31	0.030	0.000	10,916	9,883	1,033	8,760	4,370	2.3
Curitiba	10.21	0.13	9.91	0.56	0.036	0.001	5,477	5,055	422	5,919	2,527	2.0
Sao Paulo	6.24	0.09	7.00	0.29	0.030	0.000	15,630	14,435	1,195	13,465	5,935	2.4
Belem	5.48	0.36	8.18	0.62	0.033	0.000	7,258	6,145	1,113	4,358	2,469	2.5
Belo Horizonte	8.90	0.06	11.56	0.25	0.050	0.000	8,686	7,878	808	10,562	4,013	2.0
Recife	5.69	0.62	8.80	0.87	0.037	0.001	10,749	9,569	1,180	9,259	4,344	2.2
Fortaleza	6.57	0.21	12.57	0.41	0.062	0.000	10,116	9,064	1,052	8,529	3,912	2.3
Radius												
Rio de Janeiro	8.99	0.37	10.32	0.38	0.038	0.000	12,748	10,674	2,074	13,494	11,419	0.9
Salvador	7.33	1.01	10.08	1.08	0.046	0.001	10,916	8,937	1,979	8,760	7,428	1.2
Curitiba	12.06	0.73	12.60	1.33	0.048	0.002	5,477	4,765	712	5,919	5,052	0.9
Sao Paulo	7.16	0.70	9.80	0.76	0.043	0.001	15,630	13,404	2,226	13,465	11,811	1.1
Belem	6.86	1.61	9.99	2.36	0.050	0.005	7,258	5,543	1,715	4,358	3,881	1.4
Belo Horizonte	9.84	0.69	13.89	0.86	0.059	0.001	8,686	7,258	1,428	10,562	8,641	0.8
Recife	4.59	1.42	9.26	1.70	0.040	0.001	10,749	8,561	2,188	9,259	8,123	1.1
Fortaleza	7.02	1.35	14.36	2.08	0.082	0.005	10,116	8,030	2,086	8,529	7,393	1.1
Kernel												
Rio de Janeiro	7.43	0.43	9.50	0.95	0.035	0.001	12,748	11,693	1,055	13,494	12,580	0.9
Salvador	9.93	1.39	13.31	1.18	0.070	0.001	10,916	8,418	2,498	8,760	7,320	1.1
Curitiba	13.97	0.69	14.81	0.98	0.059	0.001	5,477	4,317	1,160	5,919	4,420	1.0
Sao Paulo	7.73	0.70	9.95	0.74	0.041	0.001	15,630	13,373	2,257	13,465	11,938	1.1
Belem	10.15	1.24	16.12	1.98	0.106	0.004	7,258	3,223	4,035	4,358	2,560	1.3
Belo Horizonte	9.96	0.94	12.86	0.95	0.056	0.001	8,686	7,700	986	10,562	9,242	0.8
Recife	9.65	0.44	10.78	1.10	0.041	0.001	10,749	7,090	3,659	9,259	7,344	1.0
Fortaleza	16.02	0.64	20.32	1.91	0.133	0.004	10,116	5,963	4,153	8,529	5,486	1.1

Counterfactual (3): Comparable adults living in Porto Alegre in 1998

Matching method	Absolute standardised bias				Pseudo R-squared		Treated group			Comparison group		
	Median		Mean				Observations			Observations		Average use
	Before	After	Before	After	Before	After	Before	After	Lost	Total	Used	
Nearest-neighbour												
Rio de Janeiro	8.51	0.22	12.29	0.47	0.108	0.000	12,748	11,742	1,006	11,487	4,443	2.6
Salvador	10.10	0.43	19.55	0.86	0.356	0.001	10,916	9,883	1,033	11,487	2,883	3.4
Curitiba	6.84	0.06	10.20	0.30	0.061	0.000	5,477	5,055	422	11,487	3,006	1.7
Sao Paulo	9.19	0.09	11.02	0.42	0.093	0.000	15,630	14,435	1,195	11,487	4,882	3.0
Belem	9.81	1.86	24.96	2.48	0.434	0.006	7,258	6,145	1,113	11,487	1,952	3.1
Belo Horizonte	2.47	0.13	9.21	0.42	0.135	0.000	8,686	7,878	808	11,487	3,556	2.2
Recife	14.74	0.57	23.08	1.35	0.348	0.003	10,749	9,569	1,180	11,487	2,984	3.2
Fortaleza	19.23	0.61	25.14	1.44	0.362	0.002	10,116	9,064	1,052	11,487	2,761	3.3
Radius												
Rio de Janeiro	11.88	0.32	14.24	0.46	0.103	0.000	12,748	10,657	2,091	11,487	10,117	1.1
Salvador	8.25	1.60	17.62	1.50	0.357	0.001	10,916	8,951	1,965	11,487	10,629	0.8
Curitiba	7.24	1.11	11.74	1.27	0.069	0.001	5,477	4,766	711	11,487	9,656	0.5
Sao Paulo	9.81	0.96	13.54	1.23	0.098	0.001	15,630	13,410	2,220	11,487	10,208	1.3
Belem	9.43	1.51	24.39	1.87	0.430	0.004	7,258	5,528	1,730	11,487	11,063	0.5
Belo Horizonte	4.77	0.48	10.89	0.79	0.133	0.001	8,686	7,263	1,423	11,487	10,296	0.7
Recife	10.92	1.24	22.02	1.58	0.338	0.001	10,749	8,553	2,196	11,487	10,101	0.8
Fortaleza	16.95	1.73	24.01	2.63	0.359	0.004	10,116	8,053	2,063	11,487	9,643	0.8
Kernel												
Rio de Janeiro	9.94	0.57	14.23	0.87	0.109	0.001	12,748	11,685	1,063	11,487	10,779	1.1
Salvador	8.18	1.50	15.45	1.40	0.327	0.001	10,916	8,406	2,510	11,487	10,494	0.8
Curitiba	9.70	0.63	13.07	0.87	0.066	0.001	5,477	4,352	1,125	11,487	8,779	0.5
Sao Paulo	10.26	0.78	13.07	1.20	0.085	0.001	15,630	13,320	2,310	11,487	10,286	1.3
Belem	10.58	1.28	19.84	1.30	0.311	0.001	7,258	3,163	4,095	11,487	9,386	0.3
Belo Horizonte	4.40	0.56	10.16	0.92	0.136	0.001	8,686	7,686	1,000	11,487	10,454	0.7
Recife	8.05	0.44	18.88	0.97	0.249	0.001	10,749	7,057	3,692	11,487	10,266	0.7
Fortaleza	10.44	0.88	17.97	1.96	0.235	0.003	10,116	5,949	4,167	11,487	9,734	0.6

Notes: (1) The results presented in the table refer to the sub-samples for the analysis of the dependent variable “very good or good self-assessed health”. The results of the tests for the other two health outcomes of interest (“bed due to illness” and “inability to perform usual tasks”; not shown) are virtually identical. (2) Reported pseudo R-squared of the covariates vector in the probit estimation of the “MR of residence” propensity score.

Table 6: Tests for balancing of covariates for the three preferred matching specifications, before and after matching: children

Counterfactual (1): Comparable children living in Porto Alegre in 2003

Matching method	Absolute standardised bias				Pseudo R-squared		Treated group			Comparison group		
	Median		Mean				Observations			Observations		Average use
	Before	After	Before	After	Before	After	Before	After	Lost	Total	Used	
Nearest-neighbour												
Rio de Janeiro	6.18	0.57	11.52	0.60	0.112	0.000	2,065	1,939	126	2,585	771	2.5
Salvador	13.19	0.24	24.18	0.87	0.366	0.001	2,210	1,965	245	2,585	580	3.4
Curitiba	4.14	0.23	7.80	0.71	0.043	0.001	1,125	1,039	86	2,585	602	1.7
Sao Paulo	5.90	0.18	10.20	0.85	0.067	0.001	2,986	2,752	234	2,585	888	3.1
Belem	17.89	1.39	32.04	2.56	0.432	0.006	1,672	1,448	224	2,585	467	3.1
Belo Horizonte	9.58	0.16	13.78	0.59	0.138	0.001	1,647	1,483	164	2,585	661	2.2
Recife	18.46	1.09	29.52	1.11	0.359	0.003	2,104	1,833	271	2,585	588	3.1
Fortaleza	21.50	0.67	29.29	0.99	0.341	0.001	2,303	1,982	321	2,585	619	3.2
Radius												
Rio de Janeiro	7.22	3.10	11.55	2.94	0.101	0.002	2,065	1,779	286	2,585	2,371	0.8
Salvador	12.76	3.19	22.76	3.32	0.362	0.005	2,210	1,798	412	2,585	2,327	0.8
Curitiba	4.52	3.67	7.40	3.32	0.011	0.006	1,125	983	142	2,585	2,088	0.5
Sao Paulo	5.16	1.65	10.65	2.09	0.061	0.002	2,986	2,540	446	2,585	2,308	1.1
Belem	20.87	3.14	31.40	4.11	0.428	0.005	1,672	1,298	374	2,585	2,384	0.5
Belo Horizonte	9.31	2.09	13.46	3.01	0.128	0.004	1,647	1,367	280	2,585	2,271	0.6
Recife	18.14	1.72	28.69	3.50	0.354	0.007	2,104	1,629	475	2,585	1,972	0.8
Fortaleza	18.27	3.41	28.46	5.00	0.336	0.012	2,303	1,743	560	2,585	2,104	0.8
Kernel												
Rio de Janeiro	8.09	3.32	10.62	3.21	0.073	0.002	2,065	1,554	511	2,585	2,160	0.7
Salvador	12.57	3.78	22.87	3.87	0.353	0.005	2,210	1,934	276	2,585	2,416	0.8
Curitiba	4.38	4.06	7.08	3.76	0.041	0.006	1,125	895	230	2,585	1,931	0.5
Sao Paulo	6.08	1.89	10.55	2.45	0.060	0.003	2,986	2,580	406	2,585	2,317	1.1
Belem	22.78	2.56	27.98	4.04	0.321	0.008	1,672	917	755	2,585	2,121	0.4
Belo Horizonte	8.00	1.64	12.49	3.63	0.110	0.006	1,647	1,322	325	2,585	2,200	0.6
Recife	10.53	3.16	21.19	3.79	0.212	0.005	2,104	1,298	806	2,585	2,097	0.6
Fortaleza	11.49	5.73	23.81	6.33	0.222	0.014	2,303	1,314	989	2,585	2,014	0.7

Counterfactual (2): Comparable children living in the same MR in 1998

Matching method	Absolute standardised bias				Pseudo R-squared		Treated group			Comparison group		
	Median		Mean				Observations			Observations		Average use
	Before	After	Before	After	Before	After	Before	After	Lost	Total	Used	
Nearest-neighbour												
Rio de Janeiro	6.40	0.23	6.43	0.61	0.015	0.000	2,065	1,939	126	2,412	870	2.2
Salvador	9.75	0.33	12.46	0.98	0.044	0.002	2,210	1,965	245	1,792	792	2.5
Curitiba	12.37	0.90	13.14	1.62	0.048	0.001	1,125	1,039	86	1,401	510	2.0
Sao Paulo	5.42	0.18	5.90	0.55	0.021	0.001	2,986	2,752	234	2,547	944	2.9
Belem	6.64	0.68	10.63	1.03	0.043	0.001	1,672	1,448	224	891	517	2.8
Belo Horizonte	9.59	0.16	13.05	0.36	0.047	0.000	1,647	1,483	164	2,362	726	2.0
Recife	8.65	0.46	10.02	1.25	0.022	0.002	2,104	1,833	271	2,107	806	2.3
Fortaleza	11.16	0.60	15.79	0.59	0.071	0.000	2,303	1,982	321	2,226	843	2.4
Radius												
Rio de Janeiro	6.45	1.50	6.47	2.40	0.013	0.002	2,065	1,787	278	2,412	2,118	0.8
Salvador	11.35	1.45	13.38	1.88	0.052	0.002	2,210	1,793	417	1,792	1,557	1.2
Curitiba	13.25	3.17	13.16	4.12	0.044	0.007	1,125	977	148	1,401	1,179	0.8
Sao Paulo	4.98	2.03	6.56	1.96	0.020	0.003	2,986	2,530	456	2,547	2,324	1.1
Belem	5.48	4.24	11.18	4.66	0.055	0.011	1,672	1,287	385	891	770	1.7
Belo Horizonte	9.67	4.29	13.16	5.05	0.048	0.011	1,647	1,367	280	2,362	1,885	0.7
Recife	7.52	2.07	10.52	2.94	0.026	0.004	2,104	1,624	480	2,107	1,778	0.9
Fortaleza	11.22	2.51	15.96	3.36	0.079	0.008	2,303	1,742	561	2,226	1,991	0.9
Kernel												
Rio de Janeiro	5.03	0.90	6.83	1.60	0.011	0.001	2,065	1,564	501	2,412	2,041	0.8
Salvador	11.33	1.57	12.99	2.03	0.053	0.002	2,210	1,937	273	1,792	1,652	1.2
Curitiba	15.10	4.36	13.86	4.39	0.055	0.007	1,125	889	236	1,401	1,035	0.9
Sao Paulo	5.40	1.60	6.48	1.79	0.021	0.003	2,986	2,587	399	2,547	2,349	1.1
Belem	7.85	2.83	13.20	3.43	0.065	0.008	1,672	924	748	891	644	1.4
Belo Horizonte	11.39	3.81	12.77	3.78	0.041	0.007	1,647	1,336	311	2,362	2,036	0.7
Recife	16.94	3.34	15.90	3.25	0.056	0.004	2,104	1,303	801	2,107	1,549	0.8
Fortaleza	15.22	2.29	19.55	3.59	0.108	0.009	2,303	1,334	969	2,226	1,720	0.8

Counterfactual (3): Comparable children living in Porto Alegre in 1998

Matching method		Absolute standardised bias				Pseudo R-squared		Treated group			Comparison group		
		Median		Mean				Observations			Observations		Average use
		Before	After	Before	After	Before	After	Before	After	Lost	Total	Used	
Nearest-neighbour													
	Rio de Janeiro	7.20	0.68	12.92	0.85	0.121	0.000	2,065	1,939	126	2,418	741	2.6
	Salvador	11.75	0.56	21.43	0.91	0.342	0.000	2,210	1,965	245	2,418	574	3.4
	Curitiba	10.45	0.19	12.59	0.55	0.062	0.001	1,125	1,039	86	2,418	580	1.8
	Sao Paulo	10.72	0.19	12.61	0.43	0.078	0.000	2,986	2,752	234	2,418	862	3.2
	Belem	10.15	1.07	28.63	1.53	0.439	0.003	1,672	1,448	224	2,418	437	3.3
	Belo Horizonte	7.51	0.25	14.73	0.85	0.139	0.001	1,647	1,483	164	2,418	624	2.4
	Recife	15.71	0.71	26.58	1.30	0.348	0.002	2,104	1,833	271	2,418	558	3.3
	Fortaleza	17.99	1.21	26.15	1.73	0.311	0.002	2,303	1,982	321	2,418	601	3.3
Radius													
	Rio de Janeiro	8.41	1.73	12.63	2.28	0.108	0.002	2,065	1,779	286	2,418	2,231	0.8
	Salvador	11.12	2.86	20.55	3.22	0.335	0.005	2,210	1,797	413	2,418	2,156	0.8
	Curitiba	10.06	1.20	12.15	1.81	0.033	0.002	1,125	984	141	2,418	2,048	0.5
	Sao Paulo	16.78	1.37	13.87	1.59	0.070	0.001	2,986	2,530	456	2,418	2,161	1.2
	Belem	10.71	2.52	28.30	3.70	0.436	0.008	1,672	1,292	380	2,418	2,186	0.6
	Belo Horizonte	8.01	1.50	14.31	2.39	0.126	0.002	1,647	1,367	280	2,418	2,166	0.6
	Recife	14.50	4.65	25.52	6.23	0.343	0.014	2,104	1,610	494	2,418	1,839	0.9
	Fortaleza	17.32	5.57	25.29	6.01	0.305	0.010	2,303	1,748	555	2,418	2,057	0.8
Kernel													
	Rio de Janeiro	9.69	2.92	11.63	2.88	0.079	0.003	2,065	1,559	506	2,418	2,081	0.7
	Salvador	10.71	1.93	20.45	2.71	0.328	0.004	2,210	1,936	274	2,418	2,247	0.9
	Curitiba	9.68	1.91	12.47	2.28	0.055	0.003	1,125	893	232	2,418	1,794	0.5
	Sao Paulo	15.69	0.84	13.83	1.25	0.070	0.001	2,986	2,580	406	2,418	2,178	1.2
	Belem	12.61	2.25	24.90	3.70	0.325	0.007	1,672	917	755	2,418	2,002	0.5
	Belo Horizonte	7.35	2.22	12.76	3.26	0.108	0.005	1,647	1,332	315	2,418	2,063	0.6
	Recife	7.88	2.73	17.30	3.21	0.205	0.005	2,104	1,276	828	2,418	1,990	0.6
	Fortaleza	14.56	5.19	20.77	5.29	0.194	0.010	2,303	1,318	985	2,418	1,961	0.7

Notes: See notes to Table 5.

Table 7: ATT estimations for pairwise comparisons of PSF coverage levels: adults

Panel A - Dependent variable: Very good/good self-assessed health (adults)

		ATT estimates: comparisons against Porto Alegre (no PSF)									
		(1) Difference-in-differences (DD)				(2) Propensity score matching with difference-in-differences (PSDD)					
Metropolitan region	Median PSF coverage level (1998-2003)	ATT	Standard error	P-value	Observations	Matching method	ATT	Standard error	P-value	Observations used (treatment group)	Observations used (total)
Rio de Janeiro	2.9%	-0.002	0.008	0.806	51,292	NN	0.010	0.015	0.512	11,742	26,613
						R	0.002	0.011	0.875	10,657	44,319
						K	0.000	0.011	0.967	11,685	47,942
Salvador	4.0%	**-0.030**	0.010	0.002	44,726	NN	-0.010	0.021	0.637	9,883	20,004
						R	**-0.030**	0.018	0.089	8,936	39,621
						K	**-0.032**	0.018	0.076	8,396	38,657
Curitiba	4.9%	**-0.023**	0.010	0.023	36,446	NN	-0.028	0.018	0.132	5,055	13,821
						R	**-0.033**	0.014	0.016	4,765	31,162
						K	**-0.027**	0.014	0.061	4,317	28,430
Sao Paulo	5.9%	-0.006	0.008	0.431	54,145	NN	0.000	0.015	0.981	14,435	30,643
						R	-0.007	0.010	0.497	13,399	47,795
						K	-0.006	0.011	0.570	13,299	48,005
Belem	11.4%	**-0.028**	0.012	0.018	36,666	NN	-0.002	0.030	0.940	6,145	12,605
						R	-0.032	0.027	0.244	5,528	33,817
						K	-0.007	0.028	0.806	3,122	26,358
Belo Horizonte	16.3%	-0.006	0.009	0.509	44,298	NN	0.009	0.018	0.598	7,878	19,108
						R	-0.006	0.014	0.688	7,258	38,526
						K	-0.003	0.014	0.817	7,678	39,942
Recife	22.8%	-0.001	0.010	0.916	45,058	NN	0.030	0.025	0.226	9,569	20,030
						R	0.006	0.021	0.794	8,553	38,551
						K	-0.009	0.021	0.671	7,034	36,844
Fortaleza	24.0%	0.012	0.009	0.203	43,695	NN	0.014	0.024	0.578	9,064	18,547
						R	0.029	0.019	0.137	8,030	36,365
						K	0.017	0.021	0.423	5,932	32,967

Panel B - Dependent variable: Bed due to illness (adults)

		ATT estimates: comparisons against Porto Alegre (no PSF)									
		(1) Difference-in-differences (DD)				(2) Propensity score matching with difference-in-differences (PSDD)					
Metropolitan region	Median PSF coverage level (1998-2003)	ATT	Standard error	P-value	Observations	Matching method	ATT	Standard error	P-value	Observations used (treatment group)	Observations used (total)
Rio de Janeiro	2.9%	**-0.009**	0.003	0.010	51,296	NN	**-0.015**	0.007	0.042	11,743	26,613
						R	**-0.011**	0.005	0.028	10,658	44,322
						K	**-0.011**	0.005	0.042	11,685	47,943
Salvador	4.0%	**-0.006**	0.004	0.095	44,726	NN	-0.005	0.010	0.657	9,883	20,004
						R	-0.010	0.008	0.249	8,936	39,621
						K	-0.005	0.009	0.541	8,396	38,657
Curitiba	4.9%	0.000	0.005	0.991	36,446	NN	0.003	0.009	0.698	5,055	13,821
						R	0.002	0.007	0.804	4,765	31,162
						K	0.003	0.007	0.677	4,317	28,430
Sao Paulo	5.9%	**-0.010**	0.004	0.008	54,145	NN	**-0.013**	0.007	0.071	14,435	30,642
						R	**-0.013**	0.005	0.012	13,399	47,795
						K	**-0.011**	0.005	0.028	13,299	48,005
Belem	11.4%	-0.007	0.005	0.185	36,666	NN	-0.012	0.015	0.419	6,145	12,605
						R	-0.015	0.013	0.244	5,528	33,817
						K	-0.010	0.014	0.471	3,122	26,358
Belo Horizonte	16.3%	**-0.007**	0.004	0.080	44,303	NN	-0.003	0.009	0.694	7,878	19,119
						R	-0.010	0.007	0.114	7,258	38,531
						K	-0.010	0.007	0.149	7,678	39,947
Recife	22.8%	**-0.013**	0.004	0.002	45,061	NN	-0.008	0.012	0.533	9,570	20,030
						R	-0.011	0.010	0.289	8,554	38,555
						K	-0.014	0.010	0.167	7,035	36,847
Fortaleza	24.0%	**-0.015**	0.004	0.000	43,695	NN	**-0.021**	0.012	0.072	9,064	18,547
						R	**-0.024**	0.010	0.014	8,030	36,365
						K	-0.009	0.010	0.367	5,932	32,967

Panel C - Dependent variable: Unable to perform usual tasks due to illness (adults)

		ATT estimates: comparisons against Porto Alegre (no PSF)									
		(1) Difference-in-differences (DD)				(2) Propensity score matching with difference-in-differences (PSDD)					
Metropolitan region	Median PSF coverage level (1998-2003)	ATT	Standard error	P-value	Observations	Matching method	ATT	Standard error	P-value	Observations used (treatment group)	Observations used (total)
Rio de Janeiro	2.9%	**-0.014**	0.004	0.001	51,297	NN	**-0.016**	0.009	0.071	11,743	26,613
						R	**-0.018**	0.007	0.005	10,658	44,322
						K	**-0.017**	0.007	0.009	11,685	47,943
Salvador	4.0%	**-0.009**	0.005	0.086	44,726	NN	-0.004	0.013	0.790	9,883	20,004
						R	-0.016	0.011	0.149	8,936	39,621
						K	-0.010	0.011	0.354	8,396	38,657
Curitiba	4.9%	-0.003	0.006	0.627	36,446	NN	-0.007	0.011	0.519	5,055	13,821
						R	0.000	0.008	0.996	4,765	31,162
						K	0.000	0.009	0.967	4,317	28,430
Sao Paulo	5.9%	**-0.011**	0.005	0.017	54,145	NN	-0.011	0.009	0.211	14,435	30,642
						R	**-0.011**	0.007	0.083	13,399	47,795
						K	-0.010	0.007	0.116	13,299	48,005
Belem	11.4%	-0.009	0.007	0.179	36,666	NN	**-0.034**	0.019	0.074	6,145	12,605
						R	**-0.029**	0.017	0.090	5,528	33,817
						K	-0.019	0.018	0.302	3,122	26,358
Belo Horizonte	16.3%	**-0.010**	0.005	0.049	44,304	NN	-0.005	0.011	0.651	7,879	19,114
						R	-0.012	0.009	0.152	7,259	38,532
						K	-0.012	0.009	0.152	7,679	39,948
Recife	22.8%	**-0.009**	0.005	0.078	45,061	NN	-0.015	0.015	0.327	9,570	20,030
						R	-0.014	0.013	0.278	8,554	38,555
						K	-0.021	0.013	0.102	7,035	36,847
Fortaleza	24.0%	**-0.025**	0.005	0.000	43,695	NN	**-0.035**	0.015	0.023	9,064	18,547
						R	**-0.038**	0.012	0.002	8,030	36,365
						K	**-0.026**	0.013	0.041	5,932	32,967

Notes: Matching methods are one-to-one nearest neighbour (NN), radius (R) and kernel matching (K). Standard errors for DD coefficients are robust to heteroskedasticity and clustering at the household level. Standard errors for PSDD coefficients have been derived analytically under the assumptions of homoskedasticity and independent outcomes across observations belonging to treatment and comparison groups.

Table 8: ATT estimations for pairwise comparisons of PSF coverage levels: children

Panel A - Dependent variable: Very good/good self-assessed health (children)

		ATT estimates: comparisons against Porto Alegre (no PSF)									
		(1) Difference-in-differences (DD)				(2) Propensity score matching with difference-in-differences (PSDD)					
Metropolitan region	Median PSF coverage level (1998-2003)	ATT	Standard error	P-value	Observations	Matching method	ATT	Standard error	P-value	Observations used (treatment group)	Observations used (total)
Rio de Janeiro	2.9%	0.008	0.011	0.477	9,480	NN	**0.038**	0.022	0.080	1,939	4,321
						R	**0.030**	0.015	0.047	1,779	8,499
						K	**0.033**	0.016	0.036	1,554	7,836
Salvador	4.0%	**-0.042**	0.016	0.009	9,005	NN	-0.018	0.031	0.561	1,965	3,911
						R	-0.027	0.026	0.291	1,793	7,833
						K	-0.025	0.025	0.320	1,934	8,249
Curitiba	4.9%	0.000	0.016	0.998	7,529	NN	**-0.050**	0.027	0.066	1,039	2,731
						R	-0.022	0.020	0.277	977	6,292
						K	-0.020	0.021	0.349	889	5,649
Sao Paulo	5.9%	-0.016	0.012	0.172	10,536	NN	0.009	0.022	0.689	2,752	5,446
						R	-0.007	0.015	0.643	2,530	9,323
						K	-0.006	0.015	0.699	2,580	9,424
Belem	11.4%	0.020	0.021	0.347	7,566	NN	0.056	0.042	0.179	1,448	2,869
						R	0.038	0.036	0.291	1,287	6,627
						K	0.041	0.038	0.289	917	5,684
Belo Horizonte	16.3%	0.008	0.012	0.505	9,012	NN	0.022	0.025	0.389	1,483	3,494
						R	0.016	0.018	0.367	1,367	7,689
						K	0.001	0.018	0.978	1,322	7,621
Recife	22.8%	-0.001	0.015	0.958	9,214	NN	0.033	0.036	0.352	1,833	3,785
						R	0.006	0.027	0.814	1,610	7,199
						K	-0.005	0.026	0.852	1,276	6,912
Fortaleza	24.0%	**0.029**	0.013	0.032	9,532	NN	**0.084**	0.032	0.008	1,982	4,045
						R	**0.051**	0.023	0.026	1,742	7,894
						K	**0.053**	0.023	0.024	1,314	7,009

Panel B - Dependent variable: Bed due to illness (children)

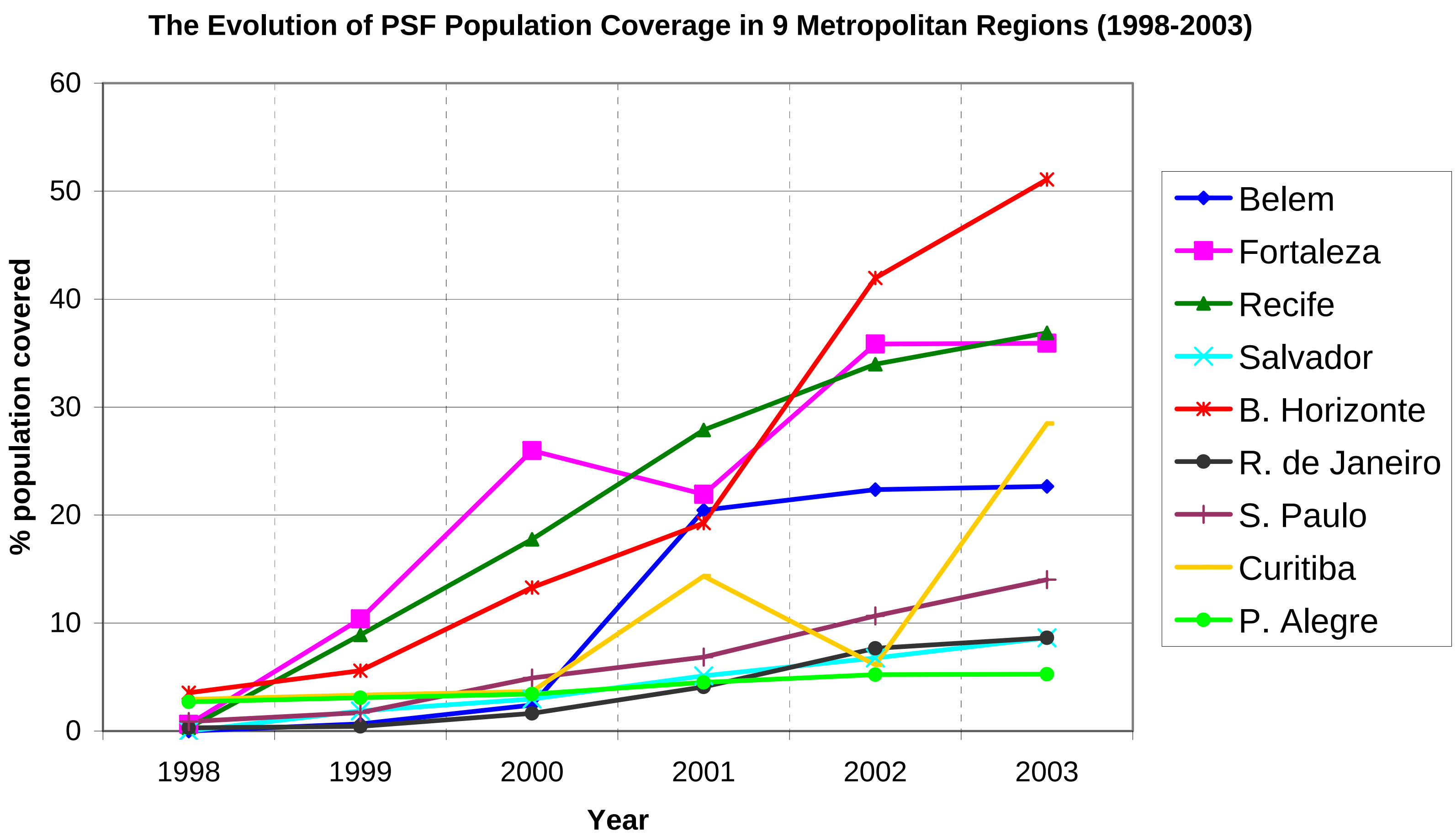
Metropolitan region	Median PSF coverage level (1998-2003)	ATT estimates: comparisons against Porto Alegre (no PSF)									
		(1) Difference-in-differences (DD)				(2) Propensity score matching with difference-in-differences (PSDD)					
		ATT	Standard error	P-value	Observations	Matching method	ATT	Standard error	P-value	Observations used (treatment group)	Observations used (total)
Rio de Janeiro	2.9%	-0.012	0.009	0.165	9,484	NN	**-0.040**	0.017	0.017	1,939	4,321
						R	**-0.027**	0.013	0.031	1,779	8,500
						K	**-0.028**	0.013	0.037	1,554	7,838
Salvador	4.0%	0.001	0.010	0.924	9,005	NN	0.001	0.023	0.965	1,965	3,911
						R	-0.010	0.018	0.586	1,793	7,833
						K	-0.001	0.018	0.947	1,934	8,249
Curitiba	4.9%	0.015	0.010	0.146	7,529	NN	-0.002	0.020	0.922	1,039	2,731
						R	0.016	0.015	0.262	977	6,292
						K	0.009	0.016	0.575	889	5,649
Sao Paulo	5.9%	-0.014	0.009	0.108	10,536	NN	**-0.029**	0.017	0.100	2,752	5,446
						R	**-0.029**	0.012	0.013	2,530	9,323
						K	**-0.031**	0.012	0.008	2,580	9,424
Belem	11.4%	0.012	0.013	0.362	7,566	NN	-0.021	0.032	0.523	1,448	2,869
						R	-0.001	0.027	0.968	1,287	6,627
						K	-0.017	0.028	0.536	917	5,684
Belo Horizonte	16.3%	**-0.015**	0.009	0.079	9,014	NN	**-0.036**	0.020	0.068	1,484	3,495
						R	**-0.026**	0.014	0.068	1,368	7,692
						K	-0.023	0.014	0.107	1,322	7,623
Recife	22.8%	**-0.015**	0.009	0.097	9,215	NN	**-0.050**	0.026	0.060	1,833	3,785
						R	-0.032	0.020	0.104	1,610	7,201
						K	-0.030	0.018	0.105	1,276	6,913
Fortaleza	24.0%	**-0.020**	0.009	0.034	9,532	NN	-0.023	0.024	0.337	1,982	4,045
						R	-0.024	0.018	0.181	1,742	7,894
						K	**-0.036**	0.018	0.047	1,314	7,009

Panel C - Dependent variable: Unable to perform usual tasks due to illness (children)

Metropolitan region	Median PSF coverage level (1998-2003)	ATT estimates: comparisons against Porto Alegre (no PSF) (1) Difference-in-differences (DD) ATT	Standard error	P-value	Observations	(2) Propensity score matching with difference-in-differences (PSDD) Matching method	ATT	Standard error	P-value	Observations used (treatment group)	Observations used (total)
Rio de Janeiro	2.9%	**-0.030**	0.010	0.003	9,488	NN	**-0.049**	0.022	0.025	1,939	4,321
						R	**-0.041**	0.015	0.008	1,779	8,504
						K	**-0.045**	0.016	0.006	1,554	7,842
Salvador	4.0%	-0.020	0.012	0.110	9,005	NN	0.034	0.029	0.245	1,965	3,911
						R	-0.013	0.024	0.574	1,793	7,833
						K	-0.011	0.024	0.658	1,934	8,249
Curitiba	4.9%	0.001	0.013	0.947	7,529	NN	-0.037	0.027	0.180	1,039	2,731
						R	-0.002	0.020	0.931	977	6,292
						K	-0.006	0.022	0.795	889	5,649
Sao Paulo	5.9%	**-0.039**	0.011	0.000	10,536	NN	**-0.063**	0.022	0.005	2,752	5,446
						R	**-0.043**	0.015	0.005	2,530	9,323
						K	**-0.042**	0.015	0.007	2,580	9,424
Belem	11.4%	0.004	0.017	0.815	7,566	NN	-0.028	0.041	0.492	1,448	2,869
						R	-0.022	0.035	0.529	1,287	6,627
						K	-0.037	0.037	0.311	917	5,684
Belo Horizonte	16.3%	**-0.028**	0.012	0.020	9,014	NN	**-0.044**	0.027	0.099	1,484	3,495
						R	-0.021	0.020	0.278	1,368	7,692
						K	-0.013	0.020	0.525	1,322	7,623
Recife	22.8%	**-0.025**	0.012	0.040	9,215	NN	-0.046	0.035	0.184	1,833	3,785
						R	**-0.052**	0.027	0.053	1,610	7,201
						K	**-0.053**	0.025	0.035	1,276	6,913
Fortaleza	24.0%	**-0.047**	0.012	0.000	9,532	NN	-0.012	0.031	0.693	1,982	4,045
						R	**-0.042**	0.024	0.072	1,742	7,894
						K	**-0.053**	0.024	0.030	1,314	7,009

Notes: Matching methods are one-to-one nearest neighbour (NN), radius (R) and kernel matching (K). Standard errors for DD coefficients are robust to heteroskedasticity and clustering at the household level. Standard errors for PSDD coefficients have been derived analytically under the assumptions of homoskedasticity and independent outcomes across observations belonging to treatment and comparison groups.

Figure 1


The Evolution of PSF Population Coverage in 9 Metropolitan Regions (1998-2003)
60
50
40
30
20
10
0
% population covered
1998
1999
2000
2001
2002
2003
Year
Belem
Fortaleza
Recife
Salvador
B. Horizonte
R. de Janeiro
S. Paulo
Curitiba
P. Alegre