

A Brief Overview on the Methods of Impact Evaluation of Public Policies

Flávio Renato Barros da Guarda*

Department of Public Health, Federal University of Pernambuco, Brazil

*Corresponding author: Flávio R B da Guarda, Department of Public Health, Federal University of Pernambuco, Brazil



ARTICLE INFO

Received: 📅 July 16, 2021

Published: 📅 August 03, 2021

Citation: Flávio Renato Barros da Guarda. A Brief Overview on the Methods of Impact Evaluation of Public Policies. Biomed J Sci & Tech Res 37(4)-2021. BJSTR. MS.ID.006047.

Keywords: Policy Impact Assessment; Physical Activity; Motor Activity; Propensity Score Matching; Difference-in-Differences Estimator

ABSTRACT

Impact assessment is an important tool for the formulation and implementation of evidence-informed public policies. The primary objective of this article is to present a brief review of the methods most commonly used in evaluating the impact of public policies, and the secondary objective is to present examples of the application of these methods in the evaluation of programs to promote physical activity. Quasi-experimental methods are configured as an important alternative for the construction of causal inferences about the impact of policies and programs, especially when it is not possible to participate in these interventions and do not happen randomly. In this sense, we highlight the propensity score matching method and the difference-in-differences estimator, which can be used alone, combined with each other or with other methods to generate valid and robust estimates of the causal effect. At the end of the article, the application of these methods in evaluating the impact of physical activity programs in Brazil and the United States is presented, emphasizing the versatility of these methods to assess the impact by comparing groups of aggregated units (such as municipalities), either to verify the effect of an intervention on groups of individuals.

Introduction

Non-communicable chronic diseases are a global public health problem, due to their high morbidity and mortality [1]. These diseases also generate temporary and permanent physical incapacities, impact the quality of life of individuals, and are responsible for high public spending on their treatment [2,3]. Among the non-pharmacological strategies for the prevention and control of NCDs, highlight the adoption of healthy behaviors, especially the regular practice of physical activity [4]. Evidence indicates that the practice of 150 to 300 minutes a week of moderate physical activity reduces the risk of developing and dying from chronic non-communicable diseases [5,6], which has led policymakers to develop strategies aimed at adopting more active lifestyles by the population. Other studies indicate that if the number of insufficiently active individuals decreased by 25%, there would be a reduction of 1.3 million deaths from CNCD each year [7]. Although the regular practice of physical activities reduces the risk of illness and death from NCDs, few studies have evaluated

the impact of public policies aimed at increasing the population's adherence to more active and healthier lifestyles.

On the other hand, the evaluation of the impact of public policies can help to demonstrate empirically to what degree a given intervention (and only it) changed the results of a variable of public interest [8]. In addition, impact assessment allows identifying the achievement of the proposed objectives for an intervention and contributes to the (re)formulation of evidence-informed policies [7]. The knowledge and use of robust impact assessment methods allow academic research to be articulated with the demands of policymakers and generate evidence that assists in decision-making processes regarding the formulation, implementation, maintenance or expansion of a policy, program or public action [8,9]. These methods are recommended by the World Bank as support tools not only to assess the effectiveness of policies, but also to verify different ways of implementing the same intervention [8]. This characteristic makes impact assessment particularly useful for

evaluating physical activity promotion programs, especially if they have different implementation strategies in the local context or if they have been proposed to serve different target audiences.

Challenges of Evaluating the Impact of Public Policies

Ideally, the impact of a public policy on physical activity should be evaluated through methods capable of comparing groups of individuals (or population aggregates) with similar characteristics, who have been randomly selected, but who differ from each other by exposure (or not) to this policy. Thus, the gold standard method for evaluating the impact of public policies is the randomized and controlled trial [9]. On the other hand, carrying out experiments with these characteristics is not always possible, both from ethical issues, as well as for financial, political, or issues related to the interest of individuals and institutions in participating in politics [10]. Personal participation or adherence of an institution, city, province, or state in a given policy does not always occur randomly [10,11], especially when it comes to behavior change related to the practice of physical activities [12]. Voluntary adherence to a policy may, therefore, follow patterns unknown by the evaluator, which limits the ability to balance the characteristics of individuals exposed and not exposed to this intervention [10,11].

In other words, self-selection for participation can result in the formation of a treatment group (group exposed to the policy) with different characteristics from the unexposed group, and this can bias the estimation of the impact of this intervention since the exposed group can be formed by those who are more motivated, with greater political will (in the case of entities, municipalities, and provinces) or who have better physical or technical conditions to adhere to physical activity [9-11]. Given the impossibility of randomly selecting participants and non-participants in a policy or program, that is, given a possible selection bias, some additional assumptions are needed to identify the parameter of interest in the impact assessment [13]. In this context, the ignobleness hypothesis describes that there will be no systematic bias when comparing groups with similar observable characteristics, as any information on heterogeneity can be captured by these observable variables, both in the group that underwent the policy and in the comparison group [14].

Therefore, the use of quasi-experimental research designs is recommended, which allow estimating the trajectory of beneficiaries of a physical activity promotion policy, if they had decided not to adhere to it [10- 12]. In this sense, the use of impact assessment methods based on counterfactual analysis is the best alternative for estimating the specific causal effects of the policy on a given indicator of interest [8-11]. The method known as the potential results model is widely used in public policy impact assessment, as it compares results of an intervention with estimates of results that

would be obtained without the intervention [8,12]. The application of the method consists of comparing (before and after) a sample of analysis units submitted to intervention (treatment group) with another sample of analysis units without intervention [8-11].

The Propensity Score Matching Method

The evaluation of the impact of public policies requires identification of what would have happened to the group exposed to a given intervention, had it not been implemented (definition of the counterfactual). However, it is not possible to construct a valid and robust counterfactual scenario by merely selecting a group of individuals who have not been exposed to the policy, as it is possible that those who are likely to be more motivated (or managers with greater commitment in terms of health policies, thus as greater political will) could be more likely to implement a physical activity program, thus characterizing a selection bias [15]. An alternative to minimize the selection bias is the use of the propensity score matching method, which allows the constitution of a comparison group with observable characteristics similar to those of the group exposed to the policy (treated) [15,16]. Consequently, it becomes possible to identify and select at least one unit from the comparison group that represents a counterfactual result for each treated unit. This scenario, therefore, creates the idea that the only difference between the units evaluated would be the participation or not in the policy, since they have other similar characteristics [16].

The Propensity Score Matching method proposes that after defining the treated and comparison groups, regression models for binary data are estimated using Logit or Probit type link functions [17] to determine the probability of an analyzed unit adhere to the policy, through a vector of characteristics from the period prior to exposure to the program ($X_{i,-1}$), which is given by:

$$P(Trat_{i,0} = 1) = \phi(\beta X_{i,-1}) \quad (1)$$

Where: $Trat_i$ is a dummy variable with a value 1 if the individual is exposed to the policy and a value of 0 for those not exposed; ϕ is an accumulated logistic distribution function, $X_{i,-1}$ is a vector of k explanatory variables weighted by the inverse of the treatment probability, and β is a vector of parameters associated with these variables.

In order to identify the best matching strategy, the next step of the method is to use the estimated propensity scores to compute weights that allow the individuals in the comparison group to be balanced, so that they become, on average, similar to the treated group. For this purpose, matching algorithms such as nearest neighbor, kernel matching and radial matching are usually used.

The Difference-in-Difference Method

The difference-in-differences estimator is a method used in quasi-experimental approaches to evaluate the impact of

interventions, based on the hypothesis of parallel evolution of the results relative to a variable of interest between the treated and comparison groups [18]. This method postulates that the trajectory of the results for the comparison group represents what would have happened with the variable result of the treated in the absence of intervention. The use of this method provides reliable estimates for the causal effect of the treatment, as long as the evolution of unobservable information from the treatment and control groups presents a uniform variation over time [18,19]. The adoption of the difference-in-differences method requires information about the treated and comparison groups before and after the implementation of the intervention that allows the construction of a scenario that describes the parallel evolution of the trajectory of the treated and the comparison group over time periods of at least one year before and after the implementation of the program. Thus, it is possible to capture the treatment effect by calculating the difference-in-differences between results observed before and after treatment [18].

Considering $t = 0$ as the period before the implementation of the intervention and $t = 1$ as the period after, the difference-in-differences estimator is described by:

$$DD_i = E[(Y_{i1}^1 - Y_{i0}^1) - (Y_{j1}^0 - Y_{j0}^0)] \quad (3)$$

Where: Y_i = result variable of a municipality treated i ; Y_j = result variable of a comparison group municipality j .

The difference-in-differences estimator is able to minimize the selection bias associated with unobservable characteristics and is therefore particularly useful to assess the impact of physical activity promotion policies, as it allows controlling problems related to time-invariant characteristics, for example, the innate abilities of individuals or the political will of a public manager to implement a program.

Combination of Public Policy Impact Assessment Methods

Importantly, impact assessments are subject to three types of bias. The first is the result of possible differences in the observable characteristics between treated and comparison groups, such as income, sociodemographic or epidemiological variables, level of training and motor experiences. The second bias may be associated with unobservable characteristics of the studied groups (such as individual motivation, political will, technical capacity of individuals, etc.). The third type of bias occurs due to the impossibility of comparing the groups due to the absence of common standards, that is, due to the absence of overlapping of the conditional density function of the observable characteristics of the treated and control group [11,16]. Biases related to observable and unobservable characteristics and the absence of a common support potentially generate imprecise conclusions regarding the impact of a public policy.

On the other hand, the control of bias in quasi-experimental studies can be performed by combining methods such as propensity score matching and regression of differences in differences, which allow to estimate the causal effect of a policy [19-21]. Impact assessment methods can be combined with each other or with other techniques, aiming to overcome possible individual limitations of the methods and increase the robustness of the results. The combination of propensity score matching and difference-in-differences methods, also known as double difference matching, improves the quality of results from non-experimental studies [20], as the difference-in-differences method minimizes possible selection biases by characteristics of the treated and comparison groups, while the matching by propensity score of units (or individuals) with similar characteristics allows to reduce both biases arising from the distribution of observable characteristics, as well as biases related to the absence of common support [21,22].

Examples of Impact Assessment of Physical Activity Programs

Physical activity promotion programs can have an impact on different variables, depending on the objectives proposed by a public policy. In this sense, in order to demonstrate this wide range of possibilities, three studies that used the impact assessment methods of physical activity programs discussed in the previous sections of this article are presented below. The studies by Rodrigues, et al. [23] and Lima, et al. [24] used the propensity score matching method to assess the impact of the Health Gym Program, which has been implemented in several Brazilian municipalities since 2011 and is considered the main health and physical activity promotion program in Brazil. The two studies used different official government databases to match municipalities according to their socioeconomic, demographic and epidemiological characteristics and to verify the impact of the program in the first state to implement this intervention in that country. The study by Rodrigues, et al. [23] evaluated the impact of the program on mortality for hypertension in the state of Pernambuco between 2010 and 2018 and found that the presence of the program promotes a 12.8% reduction in the mortality rate.

The study also revealed that the greatest effects of the program were observed in brown skin color people and in the population group over 80 years old. The study carried out by Lima and collaborators [24], also in the state of Pernambuco, verified the impact of the program on public spending on hospital admissions for cerebrovascular diseases. The results showed that the municipalities that implemented the program spent, on average, US\$ 325.22 less with hospitalizations for each group of 10,000 inhabitants than the municipalities that did not adhere to the program. In another study, Brittin and colleagues [25,26] used the difference-in-differences method to verify how much the exposure to school environments prone to movement impacted on the level

of physical and in the time on sedentary behavior activity the time of sedentary behavior of elementary school students in the New York and Virginia, in the United States. The authors observed that children who moved to schools built with the objective of decreasing the time of sedentary behavior and promoting physical activity had an average decrease of 81.2 minutes in the time of daily sedentary behavior, and an increase of 23.4 breaks in sedentary behavior and 67.7 minutes of light physical activity, when compared to students who were in traditionally built schools.

Conclusion

Public policy impact assessments aim to estimate the causal effect of an intervention designed to solve a problem of interest to society. These assessments should preferably be carried out through randomized experimental studies, but the scientific literature describes an important set of valid and robust methods that use non-experimental approaches. The use of methods that are based on potential results is an important strategy to create counterfactuals, control biases and produce reliable estimates of the impact of an intervention. Among these methods, the propensity score matching, and the difference-in-differences estimator stand out, which can be used alone, combined with each other or with other methods to create robust estimates of the causal effect of public policies when it is not possible to carry out experiments or when exposure to the intervention is not random.

Impact assessment methods are particularly useful for evaluating physical activity policies and programs, both because the participation of individuals in this type of intervention depends on personal choices, and because the promotion of more active and healthier lifestyles has become the object of several public policies aimed at mitigating the harmful effects of non-communicable chronic diseases on individuals, the economy and on expenditures in the health sector. Therefore, the need to increase knowledge and the use of impact assessment methods is highlighted in order to produce evidence on the effect of public policies. These methods can support policy (re)formulation and decision-making processes about the implementation, expansion or closure of interventions. Impact assessment methods also play an important role in generating information that allow managers to justify the investment and potential savings of public resources related to the implementation and/or adherence to physical activity programs (accountability in the public sector).

Acknowledgement

This study was funded with resources from the Research Prorectory of the Federal University of Pernambuco, Brazil.

Conflict of Interest

There is no Conflict of Interest.

References

- (2018) World Health Organization (OMS).WHO Commission Calls for Urgent Action Against Chronic Noncommunicable Diseases. Pan American Health Organization (Switzerland). Geneva.
- Rasmussen B, Sweeny K, Sheehan P (2015) Report to the Brazil-U.S. Business Council, the US Chamber of Commerce and the APEC Business Advisory Council. Victoria Institute of Strategic Economic Studies. Economic Costs of Absenteeism, Presenteeism and Early Retirement Due to Ill Health: A Focus on Brazil. Melbourne: Victoria University.
- Chaker L, Falla A, Lee SJ, Muka T, Imo D, et al. (2015) The global impact of non-communicable diseases on macro-economic productivity: a systematic review. *Eur J Epidemiol* 30(5): 357-395.
- Crowther ME, Ferguson SA, Vincent GE, Reynolds AC (2021) Non-Pharmacological Interventions to Improve Chronic Disease Risk Factors and Sleep in Shift Workers: A Systematic Review and Meta-Analysis. *Clocks Sleep* 3(1): 132-178.
- Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, et al. (2016) Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. *BMJ* 354: i3857.
- (2020) WHO. WHO guidelines on physical activity and sedentary behaviour: Web Annex. Evidence profiles. Geneva: World Health Organization.
- Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, et al. (2012) Effect of physical inactivity on major non-communicable diseases worldwide: Analysis of burden of disease and life expectancy. *The Lancet* 380(9838): 219-229.
- Gertler Paul J, Martinez Sebastian, Premand Patrick, Rawlings Laura B, Vermeersch Christel MJ (2016) Impact Evaluation in Practice, (2nd Edn.), Washington, DC: Inter-American Development Bank and World Bank.
- Bozeman B, Sarewitz D (2011) Public Value Mapping and Science Policy Evaluation. *Minerva* 49: 1-23.
- Athey Susan, Guido W Imbens (2017) The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives* 31(2): 3-32.
- Heckman J (1979) Sample selection bias as a specification error. *Econometrica* 47(1): 153-161.
- Ryan E Rhodes, Ian Janssen, Shannon SD Bredin, Darren ER Warburton, Adrian Bauman (2017) Physical activity: Health impact, prevalence, correlates and interventions, *Psychology & Health* 32(8): 942-975.
- Caliendo M, Kopenig S (2008) Some practical guidance for the implementation of propensity score matching. *Germany: IZA Discussion Papers* 22(1): 31-72.
- Firpo Sergio, Pinto Rafael de Carvalho Cayres (2013) Combining Strategies for Estimating Treatment Effects. Working Paper. São Paulo School of Economics. Center for Applied Microeconomics.
- Rosenbaum P, Rubin D (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70(1): 41-55.
- Austin PC (2011) An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Pharm Stat* 46(3): 399-424.
- Heckman J, Ichimura H, Todd P (1997) Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Review of Economic Studies* 64(4): 605-654.
- Liao T (1994) Interpreting probability models: Logit, probit, and other generalized linear models. Thousand Oaks, CA: SAGE publications 101.

19. Lechner M (2011) The estimation of Causal Effects by Difference in Difference. *Methods. Foundations and Trends in Econometrics* 4(3): 165-24.
20. Fredriksson A, Oliveira GM (2019) Impact evaluation using Difference-in-Differences. *RAUSP Manag J* 54(4): 519-532.
21. Blundel R, Dias MC (2000) Evaluation methods for non-experimental data. *Fiscal Studies* 21(4): 427-468.
22. Ravallion M (2005) Evaluating anti-poverty programs. *Policy Research Working Paper* 3625.
23. Bertrand M, Duflo E, Mullainatha S (2004) How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics* 119(1): 249-275.
24. Rodrigues BLS, Silva RN, Arruda RG, Silva PBC, Feitosa DKS, et al. (2020) Impact of the Health Gym Program on mortality from systemic arterial hypertension. *Cien Saude Colet*.
25. Lima R de CF, Rodrigues BLS, Farias SJM, Lippo BRDS, Guarda FRB (2020) Impact of the Health Gym Program on hospital admissions for cerebrovascular diseases. *Revista Brasileira de Atividade Física & Saúde* 25: 1-8.
26. Brittin J, Frerichs L, Sirard JR, Wells NM, Myers BM, et al. (2017) Impacts of active school design on school-time sedentary behavior and physical activity: A pilot natural experiment. *PLoS ONE* 12(12): e0189236.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2021.37.006047

Flávio Renato Barros da Guarda. *Biomed J Sci & Tech Res*



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>