# HIV Prevention Policy: An Econometric Evaluation

*Submitted in partial fulfilment of the requirements for the degree of*

**MASTER OF ARTS**
**IN**
**ECONOMICS**

**Rafael Aparecido Martins Frade**

*Supervisor*

**Professor Andras Danis**

CEU eTD Collection

**DEPARTMENT OF ECONOMICS**
**CENTRAL EUROPEAN UNIVERSITY**
**June 2023**

# Abstract


## HIV Prevention Policy: An Econometric Evaluation

HIV remains a social and economic challenge in many countries, despite numerous efforts to control it. A promising policy developed in the past 10 years, the Pre-Exposure Prophylaxis (PREP) have been implemented in many countries and show promising results. Most studies about PREP are either qualitative research that tries to analyse the experience of people taking it or clinical RCTs that follow small groups of people for several years. To the best of my knowledge, the present research is the first that tries to evaluate the effect of this policy as an observation study at a city or national level. The study uses data from Brazil and exploits the fact that the great majority of the people following this protocol in Brazil belong to HIV vulnerable groups. I find that, if a city has 100 new HIV cases in the vulnerable group in a given year, giving the PREP medication to 1000 people belonging to this group is likely to decrease the HIV cases in the following year by 12 to 20%.


CEU eTD Collection

# Acknowledgments


## HIV Prevention Policy: An Econometric Evaluation

First, I would like to thank my current supervisor, Andras Danis, for accepting a subject which is not traditionally studied at CEU. Also for understanding that I was working while writing the thesis, so I had limited time to work on it. Besides, for all the valuable suggestions even if health economics is not his area of expertise.

I would also like to thank CEU Econ Professors. First, for accepting someone who didn't have an economics background. Second, for the immense knowledge provided in carefully prepared classes and assignments. Third, for accepting the many extensions I needed to submit assignments. Not having an economics background, it was tough to understand all the new concepts and methods. The patience and support from the faculty were essential for me to graduate.

I would like to thank CEU and George Soros for funding my masters. Without the generous support I would have never been able to have a master's degree. Also to CEU staff, mainly Richard Kartosonto, Zsuzsanna Bordas and Corinne Freiburger who helped me in many moments of desperation.

Special thanks to Robert Lieli. He was the professor who taught me the most at CEU and the one who taught my favorite course: non-parametric econometrics. He was also my supervisor in 2021. Even though I was not able to finish my thesis then, I learned tons in the process of trying to understand the conditional average treatment effect estimator.

I don't have enough words to thank my classmates Belma, Dilnovoz, Ekaterina, Saket and Tram. Without all the help and support I received, sometimes trying to finish an assignment at 5am, I would never be able to graduate.


CEU eTD Collection

Thanks to my dear friends Bernardo, Marilia and Sara, for all the moments of joy and support. CEU would not have been the same thing without you.

Thanks to all my friends in Brazil, who I miss everyday. Special thanks to Regis for years of friendship and the valuable information on how PREP works.

Obrigado a minha mãe e minha família por tudo e mais um pouco!

CEU eTD Collection

# Abbreviations

Abbreviations used in the thesis

| | |
|---|---|
| ART | Antiretroviral therapy: medication that prevents an HIV infection to develop into Aids |
| PREP | pre-exposure prophylaxis: medication that protects against HIV infection |

CEU eTD Collection

# Contents



CEU eTD Collection

CEU eTD Collection

# List of Figures



CEU eTD Collection

# List of Tables



CEU eTD Collection

# Chapter 1

# Introduction

Despite the advances in the treatment of HIV/Aids in the past decades, the infection remains one of the most concerning issues in public health. According to the UN agency for HIV/Aids (UNAIDS), in 2021, we had the following figures:

- 38 million people living with HIV worldwide, 15% of them not aware of it,
- 1.5 million new infections every year,
- between 500 and 800 thousands deaths caused by HIV every year,
- Only 78% of the population living with HIV have access to antiretroviral therapy.

After the development of the antiretroviral therapy (ART), the main focus of the health policies has been on creating awareness and increasing the availability of prevention, making tests easy and accessible and increasing the ART coverage. These efforts, though, face difficulties from lack of financial resources to stigma and discrimination. People might avoid testing for the fear of having a positive result, and, even after a positive result, might avoid starting the ART to avoid discrimination and violence (UNAIDS). The UN agency also highlights the fact that this discrimination may come even from the health care professionals themselves. UNAIDS developed a discrimination index and, according to it, in Brazil, for example, 64% of people living with HIV have already suffered a type of discrimination due to the fact that they live with HIV (UNAIDS-Brazil).

CEU eTD Collection

A promising policy to decrease HIV infection rates is the Pre-Exposure Prophylaxis (PREP), a medical protocol that protects people following it from getting HIV even if they are exposed to the virus. As we'll see in chapter 2, many studies have proven the high effectiveness of this policy. All those studies are RCTs that follow a group of people for years, and they all converge in showing that a person following this policy is protected at a rate of virtually 100% against the infection. What is also surprising is that, to the best of my knowledge, there are yet no observational studies that try to evaluate this policy in aggregate levels, that is, how this policy could affect the general level of HIV infections in a city or a country. That's the objective of this research.

This study will try to evaluate the effect of implementing this policy in Brazil. Brazil started offering the medication to vulnerable groups in 2018 and the number of people taking it increases every year. The research question is not trying to evaluate if people taking PREP are protected, clinical studies have shown that. What I'll will try to evaluate is, given a level of treatment offered in a city or region, how much the level of new HIV infections is expected to drop.

One important aspect of the infection is that certain vulnerable groups have been historically more affected. The key populations, as defined by the UN agency for Aids, account for 95% of the infections outside sub-saharan Africa. Those are: "sex workers and their clients, gay men and other men who have sex with men, people who inject drugs and transgender people" (UNAIDS). The risk of getting HIV for those groups is 15 to 30 times higher than in the population not belonging to them. That's why, in many countries, PREP tends to be offered to these groups. I'll use the fact that vulnerable groups, mainly gay man and other man who have sex with man, are the vast majority of the people taking PREP, to establish a treatment and a control group and try to draw causal conclusions.

For this research there were several challenges. First, data on the number of people following the protocol is not easily available for researches, I had to write an involved web scrapper to obtain it. Making this data available to other researchers is one of the main contributions of this paper. Second, cities that implemented the policy had different sizes and implemented the policy in different levels, so there was a challenge on how to make them comparable. The

CEU eTD Collection

strategies taken to solve this issue can work as an example to other researchers trying to evaluate policies with similar characteristics. Third, two years after the policy started to be implemented the Covid pandemic spread around the world, which was a massive confounding factor. As we'll see, both treatment and control HIV infections decrease after that.

The structure of the paper will be: In chapter 2 I describe the results of the RCTs that evaluate the protocol and which are the other strategies that have been tried to prevent HIV. In chapter 3 I describe the data, I also discuss the empirical strategy and show how the variables were defined. Chapter 4 contains the regression results and chapter 5 concludes.

CEU eTD Collection

# Chapter 2

# HIV policies and research question

## 2.1 Strategies to prevent HIV

After the discovery of the ART, most of the efforts to decrease HIV infection rates were concentrated on increasing testing, creating campaigns on safe sex practices and making ART affordable and accessible. Many of these efforts, though, have limited effects due to lack of financial resources.

One of the strategies to create incentives for people to have safer sexual practices has been to implement cash transfer programs. Baird et al. (2012) document how a program that made yearly payments to students ranging from 1 to 5 USD and to their families ranging from 4 to 10 USD were able to reduce in 60% (even if in a small sample) the number of infections in the treatment group.

In a study published in the AEJ: Applied Economics, Björkman Nyqvist et al. (2018) studied the effect of a lottery program in Lesotho, which paid the equivalent of between 20 and 40 USD for participants who maintained a negative test result for two sexually transmitted infections (STI's). Besides the payment, the participants also had the chance of winning a large prize. The authors found that the rate of new HIV infections in the treated group was 21% lower than the control group. The authors add that traditional programs based on condom promotion or counseling on safe-sex practices were not able to reduce hiv incidence.

The evidence from cash transfer programs, though, is mixed and seems to depend on the amount of the transfer. A program in Tanzania providing cash transfers for people who main-

CEU eTD Collection

tained negative results on a set of STIs found that the treated group offered 30 USD a year didn't have a significant reduction in the incidence of STIs, while an important reduction was found for the group receiving 60 USD (De Walque et al., 2012).

## 2.2 Pre-exposure prophylaxis

Since the development of the ART, the main new scientific milestone in the fight against HIV was the development of the pre and post-exposure prophylaxis (PEP and PREP). HIV pre-exposure prophylaxis is a medical protocol that allows patients taking it to be safe from getting HIV even if exposed to the virus. In practice it consists of a single pill of Tenofovir and Emtricitabine that has to be taken everyday[1] (Brazilian Health Ministry, 2022). A week after starting the protocol, the person is already protected against HIV at rates of nearly 100%. The post-exposure prophylaxis, likewise, is a protocol that prevents the infection by the virus even if a person was already exposed to it. It is the effect of the pre-exposure prophylaxis (PREP) that is the object of analysis in this paper.

The first double-blind RCT evaluating the results of the PREP was the IPrex study, which started in 2007 and was concluded in 2011 (Grant et al., 2010). The study found a reduction of 44% in the infection rate in the treatment group and a strong correlation between the level of the medicine in blood tests and infection rates, which suggests that those infected were not correctly following the protocol. Today, a person correctly taking PREP is considered to be virtually 100% protected against the infection. Several studies conducted later found similar results: a drastic reduction in the number of HIV cases in the treatment group (Grant et al., 2014; Molina et al., 2015). Compelling evidence of the effectiveness of the protocol was given by the PARTNER study, in which 1166 serodifferent couples (when one person in the couple has HIV and the other doesn't) were given PREP and followed for 4 years in different countries and there was not a single infection in the individuals correctly following the protocol (Rodger et al., 2016). In 2022, PREP was available in more than 100 countries (Prepwatch, 2022).

[1]Some variations of this protocol are being tested (Brazilian Health Ministry, 2022).

CEU eTD Collection

## 2.3 PREP in Brazil

Since the start of the HIV epidemic, Brazil has been one of the countries that made the most significant efforts to control the spread of the infection as well as to offer treatment free of charge for people living with HIV. Within its public health system, Brazil started to offer PREP in 2018 and the number of people using it has increased significantly since then.

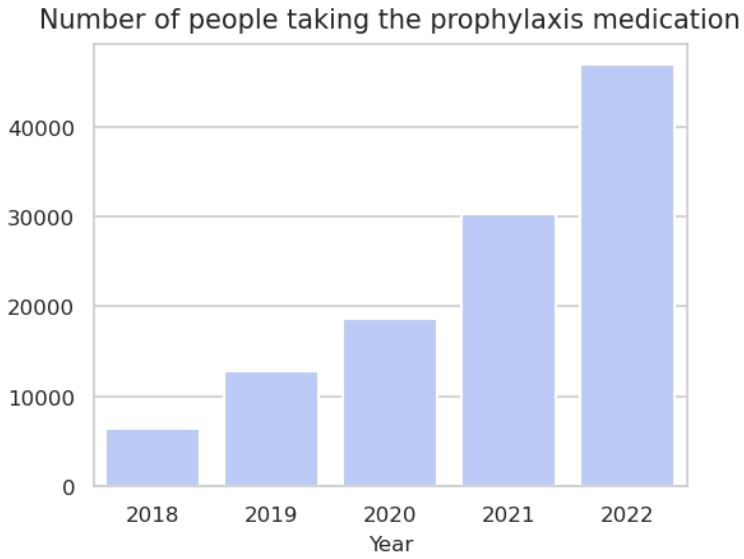


**Figure 2.1: PREP numbers in Brazil.** This graph shows the number of people following the PREP protocol at the end of every year in Brazil from 2018 to 2022.

PREP in Brazil is given to populations considered to be at an increased risk of getting HIV, like people who frequently have unprotected sex, sex workers or sero-divergent couples (Brazilian Ministry of Health, 2018). In practice, the great majority of users are gay men and other men who have sex with men (Brazilian Ministry of Health, 2022).

It is important to emphasize that the PREP medication is offered to people that do not have HIV. The whole point of the protocol is for them to be protected against the infection. For the research design, the consequence is that there is selection into treatment not only by the cities who decide to offer the treatment, but also by the individuals who decide to join the protocol. We might speculate that these individuals might have riskier behavior, that's why they would decide to take PREP, or even the opposite, they might be individuals more concerned about their health and that's why they decide to join. The assumption here is that

CEU eTD Collection

who those individuals are does not matter for the identification strategy. What I am concerned in finding is: giving a level of new HIV cases for a city, if this city offers the medication to $x$ percent of individuals defined as a proportion of the number of new HIV cases, how much is the level of new infections expected to drop. I am not trying to evaluate the efficacy of the medicine for the individuals taking it. Many RCTs have already shown that an individual taking the medication is protected at rates of nearly 100% from getting HIV. What I want to answer is: if a city has, for example, 1000 new HIV cases in 2023, which would be the number of people taking PREP that it should target if the city wanted this number to drop, let's say, 10%, in 2024?

CEU eTD Collection

# Chapter 3

# Data description and empirical strategy

## 3.1 Data Description

The data comes from two main data sources, both from the Brazilian Ministry of Health: the dataset with new HIV infections[1] and the dataset with the number of people taking PREP[2]. Both datasets have information at the city and year level. The data on HIV infections has observations since the start of the epidemic, in the 1980s and the PREP dataset has observations since the start of the PREP program in 2018 till 2022.

It is relevant to mention that, even though the PREP data is available online through a website for the general public, it was not available as a dataset, so you can search for a city and a year and see how many people are taking PREP in that city had, but the data was not consolidated in a csv or dta format. I had to write a web scraper in python to create a dataset of the number of people following the protocol in all Brazilian cities and years [3].

The HIV cases dataset has information of new HIV cases for every year and every city, and it also contains the number of new cases for several subgroups, including new cases by gender, age range, education and sexuality. I'll use this subgroup information to build the treatment and control groups and implement the identification strategy. Since the vast majority of people taking PREP were homosexual men, and there is information for HIV cases for homosexual and heterosexual men, I can consider as treated the homosexual men and as a counterfactual

[1] indicadores.aids.gov.br

[2] gov.br/aids/pt-br/assuntos/prevencao-combinada/prep-profilaxia-pre-exposicao/painel-prep

[3] The data and code used to scrape it will soon be available online github.com/rfrade/prep

CEU eTD Collection

heterosexual men. So the identification will come from the comparison between new HIV cases between these two groups. Bearing in mind that this policy has a clear start date and I have data for a control and a treatment group before and after the treatment, it is straightforward to think of difference in differences as the identification strategy.

An important characteristic of providing the PREP medication is that all cities in Brazil could have started implementing the policy since 2018, but in 2018 only a few cities were able to do so, and the number of cities has increased in the following years. So it is a clear case of selection into treatment. We might think that cities in which HIV was a bigger concern were the ones that decided to implement the policy as opposed to cities where it was not. Or even that the cities that implemented it were the ones with more infrastructure. Since I have data for the number of cases in the group of homosexual man (treated) and heterosexual man (control), for all cities, this selection into treatment should not invalidate the identification strategy. As we'll see below, the parallel trends between treated and control groups, one of the main assumptions of DiD design, shows that a city in which HIV was a concern, it is very likely that it was a concern for all groups.

Another important assumption for the identification strategy is that homosexual men being treated should affect only the new HIV cases for this group and not affect heterosexual men and women. As we could see in the start of the epidemic in the 1980s, this assumption does not hold in the long term. Even though the HIV cases were highly concentrated in the homosexual group, after a few years the heterosexual group was increasingly affected. So, even though this assumption does not hold in the long term, I do think it is likely to be valid in the first few years after the implementation of the policy.

CEU eTD Collection

### 3.1.1 Variable definitions

In this section I'll cover in detail how the treatment and outcome variables were defined. The definitions involve several assumptions which are all stated and, whenever possible, justified. It is widely known how small changes in the definition of the variables can highly impact the results. I'll try to be as clear as possible and give as many examples as I think may be necessary.

The main difficulty I found with the data is that not only cities selected into treatment,

but also cities with very different population sizes and cities that provided PREP in very different levels. For example, there are cities with a population of 50 thousand inhabitants that provided the medication to 10 persons, and cities with a population of a million that provided the medication for 100 persons. In order to make them comparable, I'll define the outcome variable as the difference in variation of new HIV cases between treatment and control groups with respect to 2018, the baseline year. For example, if city A had 100 new HIV cases in the heterosexual group (control) in 2018 and 120 in 2019, this city would have had an increase of 20% in the number of cases. If the same city had had an increase of 15% in the homosexual (treated) group, the outcome variable for 2019 for this city would be -0.05, indicating that in comparison to the control group, the treatment group had 5% less cases.

To define the treatment variable, it is important to consider that the treatment will certainly have cumulative treatment effects on the outcome variable. That means that if a city offered the treatment in 2019, even if it stopped offering it in 2020, people who would have been infected in 2019 and were not, due to the medication, would have discovered the infection in the following years. So the estimation needs to account for these cumulative effects and that's also a reason why two-way fixed effects are not suitable for the estimation (Roth et al., 2022).

The data available contains the number of people who were taking the PREP medication at the end of every year. Which does not mean that all of them started at the beginning of the year and were protected the whole year. I decided to build the variable which indicates how much treatment was offered in a city in a way that it could account for how many people were effectively protected in a given year. A quick look at the data[4] shows that people start taking the medication throughout the year and I'll use that fact to build the treatment variable. For example, if a city had 120 new people following the prophylaxis protocol in a given year, let's consider, as a first approach, that 10 started each month. So those 10 who started in January could be considered protected for the whole year, but those who started in December would be protected only for the month of December, that is, 1/12 of the year. If we compute the sum of the series, we have $(10 \cdot 12/12 + ... + 10 \cdot 1/12) = 65$. That means that if 120 new people started taking the medication in a given year, I could consider that equivalent to 65 effectively

[4]Check the online panel provided by the Health Ministry: gov.br/aids/pt-br/assuntos/prevencao-combinada/prep-profilaxia-pre-exposicao/painel-prep



CEU eTD Collection

protected for the whole year. If we extend the analysis and divide a year in $k$ periods, $n$ persons taking the medication by the end of the year would be equivalent to $\frac{n}{k}$ persons taking the medication at each period $i$, $i$ from 1 to $k$. Each of these $\frac{n}{k}$ persons who started at period $i$ would be protect for $k - i$ periods. As $k$ converges to infinity we have:

$$\lim_{k \to +\infty} \left( \frac{n}{k} \cdot \frac{k}{k} + \frac{n}{k} \cdot \frac{k-1}{k} + ... + \frac{n}{k} \cdot \frac{1}{k} \right) = \frac{n}{2}$$

All of this math to say that if 100 persons started the treatment in a year, I'll consider it equivalent to 50 persons being treated the whole year. Once these 100 persons complete this first year, they will be considered treated for the following years. It is important to highlight that the data contains only the people who started the protocol and didn't give up. To illustrate, let's see an example:

| Year | New followers (Raw data) | Effectively protected new people in the current year | Cummulative protection |
|---|---|---|---|
| 2018 | 100 | 50 | 50 |
| 2019 | 300 | 150 | 100 + 150 |
| 2020 | 500 | 250 | 400 + 250 |
| 2021 | 2000 | 1000 | 800 + 1000 |

**Table 3.1:** Example of how the treatment variable was defined.

In this example, the 100 people who started the protocol in 2018 will count as 50 in terms of cumulative effective protection for 2018, but for the following years it will count as 100, since for the following years they will be protected for the whole year. In 2019, the city had 300 new people taking the medication, which means that I'll consider it equivalent to the 150 protected people being added that year. This city had already 100 people from 2018, so for 2019 the cumulative number of people being effectively protected will be 100 from 2018 plus 150 from 2019, which means a total of 250. Observing the table it is possible to see the same procedure being applied to the following years.

The two alternatives to this approach would be considering that those who were following by the end of a year were protected the whole year, which would artificially decrease the effect

CEU eTD Collection

of the treatment variable. Or, I could also consider that the number of people who started in a given year were protected only the year after that, which would not only artificially increase the effect of the treatment variable, but it would also make me lose all the information of the treatment effect in the last year in data, 2022.

Now that the number of people effectively treated was defined, I need to transform that into a value that can be compared across cities. To do that, I'll define the treatment variable as a proportion between the people being treated and the number of new HIV cases in the year in which the policy started to be implemented, 2018. For instance, if in a city the number of new HIV cases in 2018 was 100 and in 2019 the cumulative number of people effectively treated was 150, I'll define the treatment intensity variable for this city as 1.5 . The implication is that results become easily interpretable and fairly comparable across cities. Of course, the implicit assumption here is that a city which had 100 new HIV cases in 2018 and had 150 people following the protocol is being treated with the same intensity as a city which had 10 new HIV cases in 2018 and had 15 people following the protocol. In the Figure 3.1 below we can see how the number of cities offering higher levels of treatment increases over time.

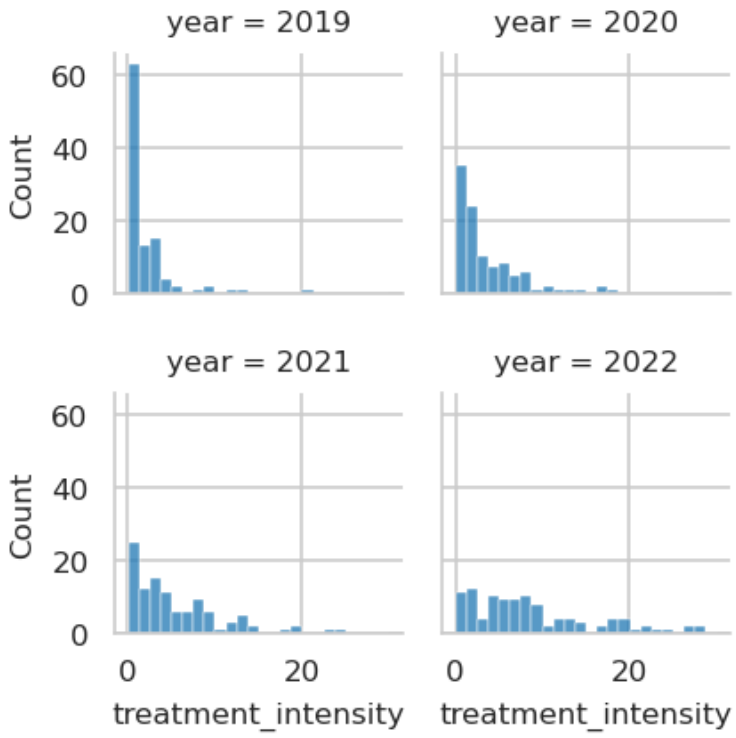


**Figure 3.1: Histogram of treatment intensity by year.** This graph shows at which level cities implemented the treatment from 2019 to 2022.

CEU eTD Collection

A consequence of defining the treatment intensity variable this way is that $n$ individuals being treated in 4 years will result in a treatment intensity equivalent to $4n$ individuals being treated in a single year. I do think that is a desirable characteristic. It allows, for example, to make cities that started in different years comparable and it also avoid the need to define variables for each year in the regression model.

Another issue for the estimation is that most cities have zero or a very small number of HIV cases and a few cities concentrate the great majority of them. The fact that a great number of cities have a small population and therefore a small number of HIV cases can introduce a lot of noise in the estimation. For instance, if city A has 1 new HIV case in 2018 and 2 persons taking PREP, the level of treatment would be equal to 2. Whereas city B with 100 new cases and offered PREP to 50 persons would have a treatment level of 0.5. We might think that city A offering PREP to only 2 persons is very unlikely to change the number of new HIV cases in the following year.

To avoid these sources of noise, I tried two solutions. First, establish some thresholds of a minimal number of cases to include the cities in the regression. Second, group cities into microregions. The Brazilian Statistics and Geography Institute has data on which cities constitute a microregion, that is, cities whose economies work around a bigger, close city. These regions would be, for example, composed of a main city where most people work, and neighboring cities where people live. This aggregation has the advantage of decreasing the level of noise in the data, but it is also plausible to imagine that if a main city has a high level of people protected against an infection, neighboring cities would also be impacted.

Bearing in mind that Brazil has 5623 cities and 510 microregions, Figure 3.2 shows how HIV cases is spread across cities and regions.

As we can see in the graph, only a handful of cities had more than 50 new HIV cases in 2018 for each group. We can also see that smaller cities tend to have less cases of HIV in the homosexual group.

CEU eTD Collection

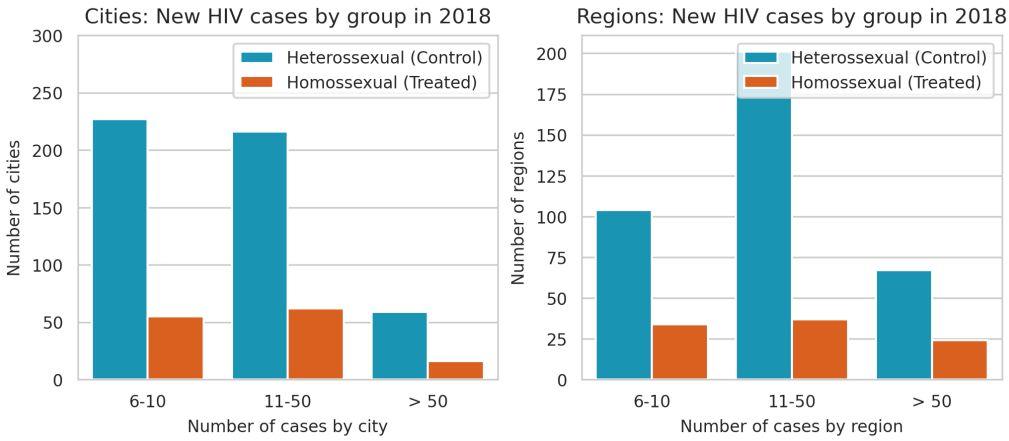


**Figure 3.2: Number of cities by number of new HIV cases.** This graph shows how many cities and regions have the number of cases in the groups in the graph, that is, from 6 to 10, 11 to 50 and more than 50.

### 3.1.2 Controls

To account for different city characteristics I'll add 3 controls to the model [5]. They try to capture characteristics that may vary over time and would not be fully captured by city or region fixed effects. Those are: city expenditure per capita in health, per capita expenditures in cash transfer programs and the percentage of population at ages between 20 and 24 years old. The expenditure in health is important to capture the local capacity to offer health care in general. The per capita expenditures in cash transfer programs is a proxy for the population living close to the poverty line, it captures the city/region general economic conditions. This is a good proxy for poverty because the payment is done by the federal government, so it does not depend on the financial conditions of the local council. That means that a higher per capita expenditure means higher levels of poverty. And the percentage of population at ages between 20 and 24 years is important since new HIV infections are more common in younger people. Cities with a high number of university students, for example, might have higher levels of new infections.

Bellow, the summary statistics with the treatment variable, the outcome variable and controls. We can see that the max of treatment intensity was 65.7, however, values like this are extreme outliers, most of the cities had a treatment intensity up to 25.

[5]Source: https://iepsdata.org.br

CEU eTD Collection

| | | (1) | | | |
|---|---|---|---|---|---|
| | n | mean | sd | min | max |
| y: difference in new HIV cases | | -.20 | .48 | -1.66 | 1.64 |
| treatment intensity | | 4.29 | 6.70 | 0 | 65.70 |
| per capita expenditure in health | | 973.55 | 456.66 | 214.92 | 3808.62 |
| per capita expenditure in cash transfers | | 59.24 | 49.76 | 4.91 | 326.89 |
| population between 20 and 24 years | | 8.10 | .93 | 5.55 | 11.34 |
| cases of covid per capita | | .073 | .068 | 0 | .36 |
| Observations | 532 | | | | |

**Table 3.2:** Summary statistics. Source of the controls: IEPS Data

## 3.2 Empirical strategy

As seen in the data description section, I'll take advantage of the fact that I have a treatment and control group for each city and the research design will be based on a difference in differences.

Given the fact that the treatment has cumulative effects in the outcome variable, a traditional two-way fixed effects is not appropriate for estimation (Roth et al., 2022). I'll exploit the fact that the treatment is continuous to estimate how a unit increase in the treatment level can reduce new HIV cases. The fact that I have a treated and a control group for each city that offered the treatment will allow me to draw causal conclusions. Stating that in terms of hypothesis:

$H_0$:a higher treatment intensity is associated with lower levels of new HIV cases.

$H_1$:a higher treatment intensity is not associated with lower levels of new HIV cases.

The first challenge when dealing with continuous treatment and staggered implementation is to plot the parallel trends. There are cities that started their treatment in 2019, others only in 2022, some with low levels of treatment intensity, others with a high level. What many researchers do is to stack the treated units at the year of implementation, so that all units can be evaluated in terms of time periods from the implementation date. In this case, however, all cities were heavily affected by covid from 2020 on, so period 1 for several cities would be the one of the covid years and a drop in HIV cases, which I could attribute to the policy, could be, in fact, caused by covid. A reasonable approach in this case would be to separate cities in two

CEU eTD Collection

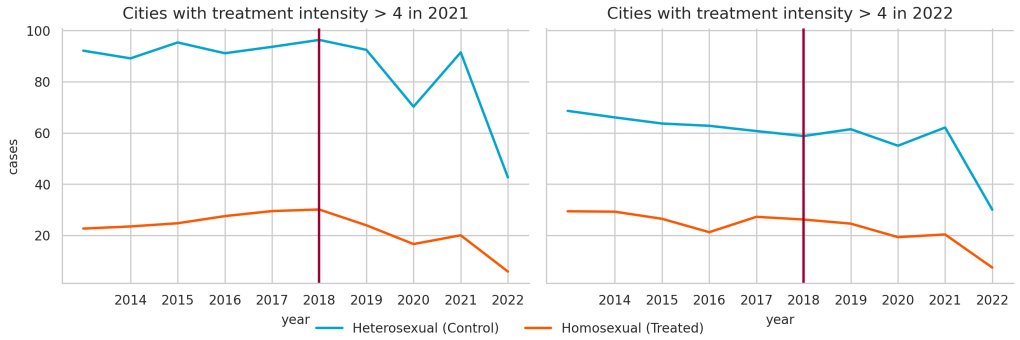


**Figure 3.3: Parallel trends.** This graph shows how the trends evolve for two groups of cities, on the left those who had already reached a treatment level of 4 in 2021. On the right, those who did it in 2022.

groups based on treatment level and plot the parallel trends, as we can see in the Figure 3.3.

In the graphs, I compared the new HIV cases in the heterosexual group (control) and in the homosexual group (treated). In the graph on the left, I selected cities that had already reached a treatment level of 4 in 2021, and on the graph on the right, cities that achieved that level only in 2022. We can see that in both cases the trends are fairly parallel before the treatment starts to be implemented. I separated them in two levels to show that the parallel trends are reasonably robust across different treatment levels. If the treatment was discrete, not continuous, I would not need to test different treatment levels, I would just need to compare treatment and control. Also, looking at the graphs, the parallel trends actually continue parallel after the first year of treatment, 2019, which could be a first sign of one of two things: the treatment didn't have any effect at all in decreasing new HIV cases, or it was not implemented in 2019 with an intensity enough to show any results. The significant drops in 2020 and afterwards in treatment and control are a direct result of covid. This research will try to identify if cases dropped more or less in cities which implemented higher levels of treatment.

The recent literature on difference in differences showed that two-way fixed effects are not appropriate to estimate a treatment with dynamic effects (Roth et al., 2022). And most of the new methods developed that deal with dynamic effects assume a binary treatment (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021). Some new methods that deal with continuous treatment are being developed, but there's still no statistical software that implements them (Callaway and Sant'Anna, 2021).

CEU eTD Collection

So, the main model will be a panel in which the outcome variable is the difference between the variation of new HIV cases in treatment and control groups and the treatment variable will be the treatment intensity, not a binary treatment indicator. That is:

$$\Delta y_{i_t} = \alpha_i + \lambda_t + \gamma X_{i_t} + \beta TI + \epsilon_{i_t}$$

Where $\Delta y_{i_t}$ is the difference between the variation of new HIV cases for treatment and control groups, both defined with respect to the baseline year; $\alpha_i$ is city fixed effects; $\lambda_t$ year fixed effects and $TI$ is treatment intensity. One of the models will also include as a control the per capita cases of covid. The panel will include cities from 2019 on if they had at least 5 new HIV cases in 2018. An alternative model will include only cities that had at least 15 new HIV cases in 2018.

CEU eTD Collection

# Chapter 4

# Results

The main hypothesis of this research is that the greater treatment intensity is, the greater would be the decrease in the level of new infections in the treatment group, in comparison to the control group. What we can see in the table below is the effect of the treatment intensity in the outcome variable, that is, the decrease in the number of new HIV cases in the homosexual group (treatment) with respect to the heterosexual group (control). City and year fixed effects were used, errors are clustered at the city level, column 3 and 4 include the cumulative number of covid cases per capita. Since the data with the controls is not yet available for 2022, columns 2 and 4 don't include 2022, resulting in less observations.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | y | y | y | y |
| treatment_intensity | -0.012** | -0.020 | -0.013** | -0.020* |
| | (0.006) | (0.012) | (0.006) | (0.012) |
| Controls | No | Yes | No | Yes |
| Covid | No | No | Yes | Yes |
| Observations | 528 | 393 | 528 | 393 |
| r2_a | 0.451 | 0.543 | 0.456 | 0.541 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 4.1: Cities with at least 5 HIV cases in 2018.** Results from regressing the variation in HIV levels between treatment and control on treatment intensity. City and year FE included, errors clustered at city level.

CEU eTD Collection

As expected, the treatment has a negative effect. A treatment of level 1 causes a decrease, on average, of 1.2% to 2% in the treated group. Assuming linear marginal treatment effects, in a city with 100 new HIV cases in a given year, having 1000 people taking PREP during a full year would be expected to decrease the number of new infections by 12 to 20% for that group. The estimates in columns 2 and 4 might not be significant as a result of having less observations.

As mentioned in the data description, there are many cities with a very small number of HIV cases and these cities can generate some noise in the results. To deal with that, the table below shows the same regressions, but only with cities that in 2018 had at least 15 new HIV cases in the treatment group.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | y | y | y | y |
| treatment_intensity | -0.006 | -0.004 | -0.007 | -0.009 |
| | (0.004) | (0.017) | (0.004) | (0.016) |
| Controls | No | Yes | No | Yes |
| Covid | No | No | Yes | Yes |
| Observations | 240 | 112 | 240 | 112 |
| r2_a | 0.397 | 0.337 | 0.394 | 0.349 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 4.2: Cities with at least 15 HIV cases in 2018.** Results from regressing the variation in HIV levels between treatment and control on treatment intensity. City and year FE included, errors clustered at city level.

CEU eTD Collection

Table 4.2 shows one of the main limitations of this research: the small number of cities with a high number of HIV cases. This estimation naturally had less observations and it might be the reason why none of the coefficients were significant.

Another way of trying to deal with the noise caused by small cities was grouping the data in regions. We can see in Table 4.3 the results of running the regression with this aggregation. The results are reasonably similar to the regressions in which the unit of observation is the cities. Assuming the number of covid cases per capita is a good proxy for the effect of covid in different regions, we can see that the results are fairly robust. Again, the difference in columns

2 and 4 might be due to a lack of data.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | y | y | y | y |
| treatment_intensity | -0.019*** | -0.002 | -0.018*** | -0.002 |
| | (0.007) | (0.029) | (0.007) | (0.029) |
| Controls | No | Yes | No | Yes |
| Covid | No | No | Yes | Yes |
| Observations | 376 | 168 | 376 | 168 |
| r2_a | 0.448 | 0.327 | 0.447 | 0.318 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

**Table 4.3: Regions with at least 5 HIV cases in 2018.** Results from regressing the variation in HIV levels between treatment and control on treatment intensity. Region and year FE included, errors clustered at region level.

Having estimated the ATT, an important extension of the analysis would be to visualize the marginal treatment effects. As we can see in Figure 4.1, there's no clear increase of the effect as the treatment intensity increases. Most of the data points are bellow zero, which suggested that, at least, the treatment didn't have the opposite effect of what would be expected. The other pattern that seems clear is that, for higher treatment intensities, almost no data points are above zero, that is, the treated group had a greater reduction in the number of new HIV cases compared to the control group.

The significant coefficients suggest that every 100% increase in the number of people taking PREP in a whole year, defined as a proportion to the number of cases, is expected to reduce the number of new HIV infections between 1.2 to 2%. Figure 4.1 shows that there might be a lot of noise in the results.

A very important policy implication is that the results are a consequence of how the policy was implemented in Brazil, where the vulnerable population to HIV are gay man and other man who have sex with man, sex workers and transsexual people. Countries where the population vulnerable to HIV have different characteristics will certainly benefit from the policy, but might find different elasticities between the treatment levels offered and the decrease in the HIV cases. Knowledge on how to target the public can be crucial in those cases.

CEU eTD Collection

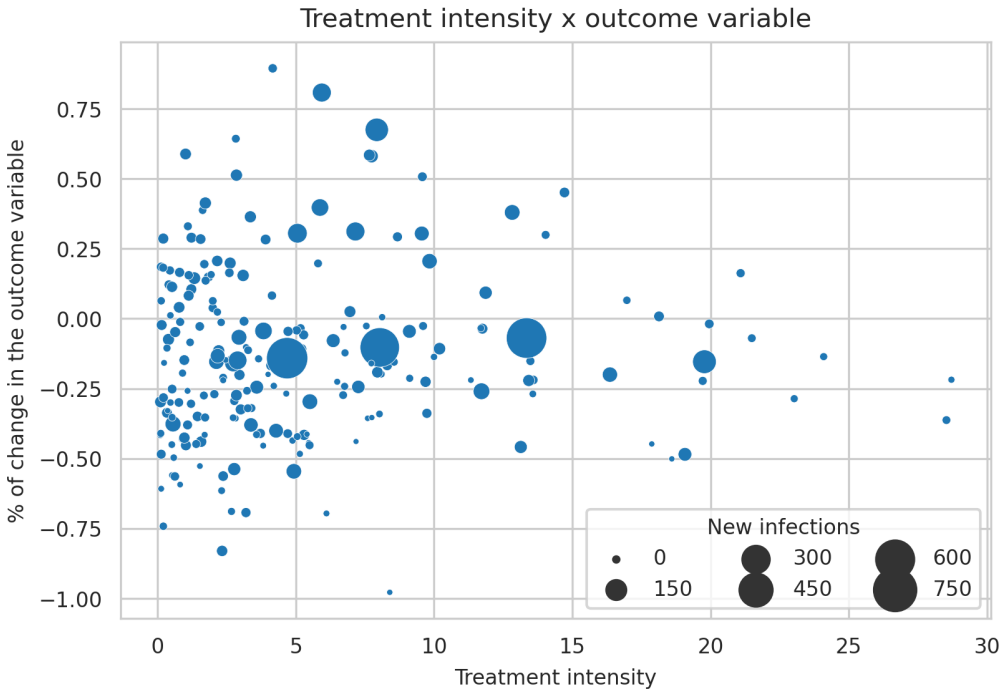


**Figure 4.1: Scatter plot Treatment intensity x outcome variable.** Size of the dots proportional to the number of infections. Negative levels in the y-axis indicate higher treatment effects, that is, cities in which the treatment group had a higher decrease of new HIV cases in comparison to the control group.

CEU eTD Collection

# Chapter 5

# Conclusions

In this paper I tried to identify the effects of the implementation of PREP in Brazil. Even though the results do suggest that the policy work as expected and it is able to reduce the overall levels of new HIV infections, data limitations and the fact that covid heavily affected the results mean that further research is necessary to provide clearer answers. Running the same regressions again with data from 2023 and 2024 will certainly provide more robust results. Besides the results themselves, this research did several other contributions to the literature.

First, for those trying to evaluate the effects of a continuous treatment, be it regarding PREP or any other health policy, this research highlights some of the difficulties of establishing which groups can be considered treated and when treatment starts. As shown, transforming the treatment effect of each period into a variable that can be interpreted as cumulative can be a solution. The details of how that is done matter a great deal. It certainly might involve several assumptions and those assumptions should be clearly stated.

Besides, one of the issues of trying to identify the effects of PREP is that cities self-select into treatment. As we know, that's not necessarily an issue for the estimation with diff-in-diff, due to the parallel trends assumption. However, when the treatment is continuous, self selection might result in few or very few units with high levels of treatment, and many with very low levels, up to a point that we might not have enough units that can actually be considered treated. As I showed, aggregating several units into larger groups can be a solution. For this aggregation to make sense, we might also need extra assumptions.

I would say that the main contributions, though, are the dataset and the descriptive anal-

CEU eTD Collection

yses. The Brazilian Ministry of Health provides the data with HIV cases in a fairly handy way. The data on people taking PREP, on the other hand, even though publicly available for visualization, is not available in csv or excel format. After a considerable effort, the data now is available to other researchers. And, besides the data, the contribution to other researchers and policy makers are the detailed description of how the policy was implemented in Brazil.

Last, it is a known fact that in statistical research, details in the estimation procedures can significantly change the results and more and more often researchers are required to provide the code that generate the results. I wanted to mention that all the code use for this research, together with the final dataset are available and the results are easily reproducible [1].

[1] github.com/rfrade/prep

CEU eTD Collection

# Bibliography

Sarah J Baird, Richard S Garfein, Craig T McIntosh, and Berk Özler. Effect of a cash transfer programme for schooling on prevalence of hiv and herpes simplex type 2 in malawi: a cluster randomised trial. *The Lancet*, 379(9823):1320–1329, 2012.

Martina Björkman Nyqvist, Lucia Corno, Damien De Walque, and Jakob Svensson. Incentivizing safer sexual behavior: evidence from a lottery experiment on hiv prevention. *American Economic Journal: Applied Economics*, 10(3):287–314, 2018.

Brazilian Health Ministry. Portaria sctie/ms nº 90, de 25 de agosto de 2022, 2022. URL `https://www.gov.br/conitec/pt-br/midias/protocolos/20220902_PCDTPrEP.pdf`. Accessed: 2023-06-01.

Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

Clément De Chaisemartin and Xavier d'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020.

Damien De Walque, William H Dow, Rose Nathan, Ramadhani Abdul, Faraji Abilahi, Erick Gong, Zachary Isdahl, Julian Jamison, Boniphace Jullu, Suneeta Krishnan, et al. Incentivising safe sex: a randomised trial of conditional cash transfers for hiv and sexually transmitted infection prevention in rural tanzania. *BMJ open*, 2(1):e000747, 2012.

Robert M Grant, Javier R Lama, Peter L Anderson, Vanessa McMahan, Albert Y Liu, Lorena Vargas, Pedro Goicochea, Martín Casapía, Juan Vicente Guanira-Carranza, Maria E Ramirez-Cardich, et al. Preexposure chemoprophylaxis for hiv prevention in men who have sex with men. *New England Journal of Medicine*, 363(27):2587–2599, 2010.

CEU eTD Collection

Robert M Grant, Peter L Anderson, Vanessa McMahan, Albert Liu, K Rivet Amico, Megha Mehrotra, Sybil Hosek, Carlos Mosquera, Martin Casapia, Orlando Montoya, et al. An observational study of preexposure prophylaxis uptake, sexual practices, and hiv incidence among men and transgender women who have sex with men. *The Lancet. Infectious diseases*, 14(9):820, 2014.

Jean-Michel Molina, Catherine Capitant, Bruno Spire, Gilles Pialoux, Christian Chidiac, I Charreau, and JF Delfraissy. On demand prep with oral tdf-ftc in msm: results of the anrs ipergay trial. In *Conference on retroviruses and opportunistic infections*, volume 2015, 2015.

Prepwatch, 2022. URL `https://www.prepwatch.org/`. Accessed: 2023-06-01.

Alison J Rodger, Valentina Cambiano, Tina Bruun, Pietro Vernazza, Simon Collins, Jan Van Lunzen, Giulio Maria Corbelli, Vicente Estrada, Anna Maria Geretti, Apostolos Beloukas, et al. Sexual activity without condoms and risk of hiv transmission in serodifferent couples when the hiv-positive partner is using suppressive antiretroviral therapy. *Jama*, 316 (2):171–181, 2016.

J Roth, PHC Sant'Anna, A Bilinski, and J Poee. What's trending in difference-in-difference? a synthesis of the recent econometrics literature. arxiv, 2022.

UNAIDS. United nations aids/hiv agency. URL `https://unaids.org/en/resources/fact-sheet`. Accessed: 2023-06-01.

UNAIDS-Brazil. United nations aids/hiv agency, brazilian office. URL `https://unaids.org.br`. Accessed: 2023-06-01.

%addtocontentstoc

CEU eTD Collection