Department of Economics, University of Warwick

Monash Business School, Monash University

as part of

Monash Warwick Alliance

How does Monitoring and Evaluation affect Racial Health Inequality? Evidence from PMAQ Program in Brazil

Taoshan Chen

Warwick-Monash Economics Student Papers

March 2023 No: 2023/51

ISSN 2754-3129 (Online)

https://warwick.ac.uk/fac/soc/economics/research/wmesp/

The Warwick Monash Economics Student Papers (WM-ESP) gather the best Undergraduate and Masters dissertations by Economics students from the University of Warwick and Monash University. This bi-annual paper series showcases research undertaken by our students on a varied range of topics. Papers range in length from 5,000 to 8,000 words depending on whether the student is an undergraduate or postgraduate, and the university they attend. The papers included in the series are carefully selected based on their quality and originality. WM-ESP aims to disseminate research in Economics as well as acknowledge the students for their exemplary work, contributing to the research environment in both departments.

"*We are very happy to introduce the Warwick Monash Economics Student Papers (WM-ESP). The Department of Economics of the University of Warwick and the Economics Department at Monash University are very proud of their long history of collaboration with international partner universities, and the Monash Warwick Alliance reflects the belief in both Universities that the future will rely on strong links between peer Universities, reflected in faculty, student, and research linkages. This paper series reflects the first step in allowing our Undergraduate, Honours, and Masters students to learn from and interact with peers within the Alliance.*"

Ben Lockwood (Head of the Department of Economics, University of Warwick) and Michael Ward (Head of the Department of Economics, Monash University)

Recommended citation: Chen, T. (2023). How does Monitoring and Evaluation affect Racial Health Inequality? Evidence from PMAQ Program in Brazil. *Warwick Monash Economics Student Papers* 2023/51.

[1] Warwick Economics would like to thank Lory Barile, Gianna Boero, and Caroline Elliott for their contributions towards the selection process.

How does Monitoring and Evaluation Affect Racial Health Inequality? Evidence from PMAQ Program in Brazil

Taoshan Chen*

Abstract

This study provides novel evidence on how monitoring and evaluation affects racial health inequality, with data from the Primary Care Access and Quality (PMAQ) Program in Brazil. By using the heterogeneity-robust estimator from de Chaisemartin and D'Haultfoeuille (2022), this study considers the non-staggered and non-binary characteristics of the treatment. The results show that an increase in monitoring and evaluation intensity can reduce racial health inequality, achieved by improving the health conditions for non-white individuals and deterioration of the health conditions for white individuals. It is suggested for policy makers to increase the allocation of health resources to ensure that while racial health inequality is reduced, both white and non-white individuals can benefit from an improvement in primary health care, rather than narrowing the gap by reducing the quality of care for one group.

JEL Classifications: I14, I18

Keywords: Racial Health Inequality, Monitoring and Evaluation, Primary Care Access and Quality (PMAQ) Program

---

* Contact Email: ctaos0101@gmail.com.
Acknowledgement: I would like to first express my sincere appreciation and gratitude to my supervisor Professor Bhalotra for encouraging me, patiently answering my questions and providing insightful comments. I also own my heartfelt gratitude to Dr. Letícia, for her generous sharing of her time and illuminating suggestions. Finally, I want to extend my thanks to Miss Beatriz, for all her kindness and help.

1. Introduction

For at least two decades, economists have been giving close attention to the subject of inequality. Race, as one of the predetermined factors, is a significant focus of inequality studies, especially in countries with a multicultural population, for example, Brazil and the United States. The conflicts from racial inequality, as with other types of inequality (e.g., gender, socioeconomic status), can trigger populism and societal stress, which may in turn harm social stability and economic development (Alesina and Rodrik, 1994; Alesina et al., 2018). Brazil is a racially diverse country with entrenched racial stratification. Different from countries with only one racial classification system that the racial record from institutions and self-reported classification are consistent (e.g., by the ethnic and ancestral in United Kingdom and United States), Brazil has two racial classification systems (Travassos and Williams, 2004; Hone et al., 2017). While institutions rely on skin color for racial classification, citizens can self-classify or reclassify as another race based on their socioeconomic status and their self-identification (Golash-Boza, 2010). For instance, black-born individuals can self-report as white in the census, if they have experienced a boost in social status. This racial mobility in Brazil has been confirmed in the literature (e.g., Carvalho et al., 2004; Schwartzman, 2007). In Brazil, there are five race categories: white (branco), black (preto), brown (pardo/mixed), Asian (amarello) and indigenous. According to the self-report data from the Brazilian censuses of 2000 and 2010, approximately 99% of Brazilians selected into one of the first three categories (i.e., white/black/brown). Although the literature on racial mobility before 2000 (e.g., Wood and Carvalho, 1994) documented that the transformation from black to brown or white was trending more than in the opposite direction, the recent census data shew an increasing proportion of residents identifying as black or brown (44.7% in 2000 and 50.7% in 2010) while the population self-declaring to be white decreased from 53.7% in 2000 to 47.7% in 2010. Most of the change was in a higher proportion of brown people (from 38.5% in 2000 to 43.1% in 2010). This trend may partially be explained by the stronger barriers to education and socioeconomic transition after 2000, especially for groups wanting to cross the educational and socioeconomic hurdle to identify as white, as the transition is much more difficult to be realized at higher level of education and social status (Hasenbalg and Silva, 2005; Ribeiro, 2006). Meanwhile, the change of education level or socioeconomic status perceived as sufficient to transition from black to brown is easier to achieve. Another reason can be the change in social norms and racial attitudes after 2000, which may result from the cumulative impacts of the Black movements in Brazil during the twentieth century, for example, the establishment of Movimento Negro Unificado (i.e., MNU) in 1978 and events thereafter (Covin, 2006).

As in other racially diverse countries, Brazil experienced many dimensions of racial inequality. For example, black and brown citizens are more likely to live below the poverty line (Lima and Prates, 2018), face unequal opportunity in employment and education, and have lower average income (Ribeiro, 2006). Apart from the socioeconomic aspects, researchers also found evidence of racial health inequality in Brazil. For example, the mortality rate for children with a non-white mother is much higher than for children with a white mother (Matijasevich et al., 2008). Non-white individuals were also found to use fewer healthcare services, such as breast cancer screenings, than their white contemporaries (Paixão, 2010). To reduce health inequality, Brazil created the Unified Health System (i.e., SUS) in 1990 to replace the previous corporative system, aiming to expand primary care coverage and to provide a free, universal, integral and equitable healthcare service with a community-based approach (Macinko and Harris, 2015; Castro et al., 2019; Ministry of Health, 2022). After primary healthcare was introduced in 1994, three channels have been created for the main provision of primary healthcare: The family health team (i.e., FHT), which includes a physician, a nurse, a nurse assistant, and 4-6 full-time community health agents; the oral health team (i.e., OHT), which consists of at least one dentist and one dentist assistant; the Family Health Support Unit (i.e., NASF), which comprises mental health and rehabilitation staff, advisors for nutrition, maternal and childcare, and pharmacy and social assistance workers (Russo et al., 2021).

The quality of the healthcare service plays an important role in increasing the efficiency of the healthcare system. Worldwide, pay-for-performance (i.e., P4P) is a popular scheme to enhance the quality of primary healthcare (i.e., PHC) (e.g., Petersen, 2007; Buetow, 2008; Kovacs et al., 2020). In 2011, Brazil implemented its first P4P program, the Program for Improving Access and Quality of Primary Care (i.e., PMAQ), in which unit of participation is the family health team. There are two main mandatory PMAQ interventions designed to improve PHC quality and access: monitoring and evaluation, and financial incentives. The funds awarded to each FHT depend on the PMAQ score, which is measured with the data collected during the monitoring and evaluation. The PMAQ program was rolled out in three rounds from November 2011 to December 2019 and it is one of the largest P4P programs in the world, with the expenditure of US$1.5 billion (R$8.6 billion) (FNS, 2022). In the last five years, the literature on the PMAQ program bloomed and most of it investigated the effect of the PMAQ program as an integrity (e.g., de Medeiros et al., 2020; Russo et al., 2021), or focused on the quality of PHC and financial incentives (e.g., Kovacs et al., 2021; Fardousi et al., 2022). However, there is scant evidence of the effect of monitoring and evaluation, especially on racial health inequality. To address this shortfall, this study applies the difference-in-differences identification strategy to evaluate how monitoring and evaluation affects racial health inequality in Brazil, using the PMAQ program setting.

There are three possible paths through which the intervention of monitoring and evaluation in the PMAQ program can affect racial health inequality. The first is the incentive for staff to improve their performance, as measured by the PMAQ indicators. During the monitoring process, data on service utilization is collected and the external evaluation indicators are classified into three modules: the Basic Health Unit (i.e., UBS); the users' satisfaction, and the family health team's work process. Since the performance indicators depend on the outcomes for white and non-white individuals, the UBSs and FHTs are encouraged to serve both racial groups better. This is an incentive built into the monitoring and evaluation system that should reduce racial discrimination in service delivery, and thus results in lower racial health inequality. Secondly, monitoring and evaluation can contribute to reducing racial health inequality by increasing the efficiency of the primary healthcare system. Despite the declaration of equal service delivery by the Unified National Health System, several researchers were skeptical about it (e.g., Matijasevich et al., 2010; Paixão, 2010; Victora et al., 2011; Chor, 2013). In a system with biased service delivery, a shortage of human or physical resources may be one critical contributor to the persistence of inequality. For instance, according to the indicator of the FHT population coverage collected in the external evaluation process, the population covered by one FHT can range from less than 1000 (e.g., in Porto Velho and Boa Vista) to more than 10000 (e.g., in Campinas and Taboão da Serra). Inadequate resources would exacerbate the racial health inequality if the allocation of limited resources were biased in favor of white people. In this situation, higher quality and more efficient service delivery brought about by monitoring and evaluation (Unger and De Paepe, 2019) could alleviate the scarcity of health resources, such as hospital beds and staff, leaving more resources available for non-white patients, which could in turn reduce racial health inequality even if racial discrimination exists. The third channel originates in the self-adjustments of behavior induced by the Hawthorne effect (Landsberger, 1958). Under monitoring and evaluation, physicians and other service-delivery staff are assumed to be more cautious. In a health system that is pursuing equality of treatment, staff are assumed to limit their racially-discriminating behavior when they are being monitored and evaluated.

By exploring the effect of monitoring and evaluation on racial health inequality in the PMAQ setting, this study contributes to several streams of literature. Firstly, this study enriches the research about the racial health inequality in Brazil and sheds light on whether monitoring and evaluation reduces racial health inequality. Secondly, it expands the literature related to the role of monitoring and evaluation in P4P strategy, especially the PMAQ program in Brazil. Thirdly, this study provides valuable information for government and policy makers to develop and modify policies related to monitoring and evaluation process, and racial health inequality.

The rest of the research is constructed as follows. In Section 2, I describe the PMAQ Program; in Section 3, I outline the empirical strategy; in Section 4, I present my study sample; in Section 5, I describe the data source and summarize the descriptive statistics; in Section 6, I present the empirical results; in Section 7, I show the robustness checks; and in Section 8, I draw a conclusion, and discuss the implications and limitations of this study.

2. PMAQ Program

Universal and equal access to primary healthcare has been an important goal for the Brazilian Ministry of Health since the establishment of SUS in 1990. After reforming and improving primary healthcare delivery for about twenty years (Facchini et al., 2018), in 2011, Brazil launched its first P4P program, the Program for Improving Access and Quality of Primary Care (i.e., PMAQ), which was implemented in 3 rounds from 2011 to 2019. The first round was from November 2011 to March 2013; the second round was from April 2013 to September 2015 and the third round was from October 2015 to December 2019. The PMAQ was implemented at the federal level and the participation unit is the family health team (i.e., FHT). The FHTs are deployed based on a community approach and each FHT is responsible for up to 1000 households (Macinko and Harris, 2015). There are four main PMAQ interventions. Two are compulsory: Monitoring and Evaluation, and Financial Incentives, and two are voluntary but encouraged: Continuing Education and Institutional Support. At the beginning of each round, a PMAQ score is calculated based on the assessment of the performance of each FHT. The assessment includes the internal self-assessment, routine monitoring, and independent external evaluation conducted by the universities. Although the last two interventions are not mandatory, the performance indicators related to these two parts will be included in the self-assessment and external evaluation. The self-assessment is voluntary and once the teams submit their self-assessment, they will receive full marks for this part regardless of their answers (Kovacs et al., 2021). The indicators for monitoring process fall into seven parts: 1) women's health (especially pregnant women); 2) children's health; 3) diabetes mellitus and hypertension; 4) oral health; 5) overall production; 6) tuberculosis and leprosy; and 7) mental health. In the external evaluation, there are three modules: 1) condition of the infrastructure, materials, supplies and medicines of the Basic Health Unit; 2) the efficiency of teams' working processes and systems; and 3) users' satisfaction and perception of access and use of health services. Information about the indicators can be found in the instruction manuals from the Ministry of Health[2].

[2] Instruction Manual for the first round: http://189.28.128.100/dab/docs/publicacoes/geral/manual_instrutivo_pmaq_site.pdf
Second round: http://189.28.128.100/dab/docs/portaldab//publicacoes/manual_instrutivo_PMAQ_AB2013.pdf

According to the PMAQ scores, the performance of the FHT will be classified into four groups (five in the third round) (i.e., worst/worse/(middle)/better/best) and there will be monthly financial incentives for the FHTs based on the performance group they are placed in. During the first two rounds, the PMAQ score for each FHT is adjusted by the socioeconomic inequality outcome. In the third round, the classification process depends on the pure PMAQ score. The instantaneous financial award will be transferred to the municipalities that the FHTs belong to once they register the program and the PMAQ scores are given. Although PMAQ funds are required to be spent on health, municipalities have the autonomy over the allocation of this income. Additionally, municipalities can decide whether to register the program and which team to enroll in the PMAQ program. The PMAQ participation is voluntary and the participation rate is increasing over the period. The details of the three rounds of the PMAQ program and the participation of municipalities and FHTs are exhibited in Table 1.

Table 1 Number of Municipalities and Family Health Teams that Participated in Each Round

Round	Num of Registered Municipalities	Num of Registered FHTs
Round 1: Nov 2011-March 2013	3965	17482
Round 2: April 2013-Sep 2015	5072	30521
Round 3: Oct 2015-Dec 2019	5324	38865

Third round: http://189.28.128.100/dab/docs/portaldab/documentos/Manual_Instrutivo_3_Ciclo_PMAQ.pdf

3. Empirical Strategy

In this study, I use difference-in-differences as the identification strategy. However, recent literature has cast doubt on the robustness of TWFE estimates in classical difference-in-differences and event study when there is a heterogeneous effect either within the treatment group or across the groups with different treatment timings (e.g., Boruskay et al., 2021; Goodman-Bacon, 2021; Athey and Imbens, 2022). The bias is caused by the negative weights of some heterogeneous effects when calculating the TWFE estimates as a weighted average of the treatment effects. Besides, in a setting where treatment is not binary but has different intensities, the TWFE estimates can also be contaminated by selection bias, especially when measuring the ATT (Callaway et al., 2021). As mentioned in Section 2, municipalities can decide whether to register the PMAQ program and how many FHTs under their jurisdiction will participate in each round. In this case, the number of the FHTs for a specific municipality may vary during different rounds. Therefore, the treatment is non-staggered (i.e., non-fixed treatment intensity)[3] and non-binary in the setting of this study and the heterogeneous treatment effect is plausible to exist, which implies a poor TWFE estimator performance with the simple difference-in-differences and event study. To date, many alternative heterogeneity-robust estimators have been proposed (e.g., Borusyak et al., 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Wooldridge, 2021; de Chaisemartin and D'Haultfoeuille, 2022). Considering the non-staggered and non-binary treatment in the PMAQ setting, this study opts for the one proposed by de Chaisemartin and D'Haultfoeuille (2022), which

[3] The non-staggered treatment here implies that the treatment intensity can change after the unites get treated, which is slightly different from its traditional explanation (i.e., treatments with multiple time periods). The definition can be found from the literature related to the advanced heterogeneity-robust estimators, for example, Callaway and Sant'Anna (2021) and de Chaisemartin and D'Haultfoeuille (2022).

also allows for time-varying covariates and dynamic effects from the past treatment, in addition to the non-staggered and non-binary treatment. Since the options for the non-binary treatment are limited to the discrete treatment with non-stagger setting and continuous treatment with staggered setting in de Chaisemartin and D'Haultfoeuille (2022), the former will be employed here, due to the flexibility of participation intensity for municipalities during the 3 roll-out cycles in PMAQ. The details of derivation can be found in the Web Appendix Section 2 of de Chaisemartin and D'Haultfoeuille (2022). The estimates can be interpreted as the average total effect (sum of the instantaneous and dynamic effect) per unit intensity of treatment and are measured as the intention-to-treat effect on the outcome from the first treatment, scaled by the first-stage effect of the first treatment on the subsequent treatment, with following equation:

$$\delta = \frac{\sum_{f=0}^{Fu} w_f DiD_t}{\sum_{f=0}^{Fu} w_f \delta_f^D} \tag{1}$$

where F_u represents the difference between the last period when the never-treated group exists and the first period when a group becomes treated; w_f is calculated as $N_f^1 \ / \ \sum_{f'=0}^{F_u} N_{f'}^1$, where N_f^1 is number of switchers in groups reaching f periods after the first treatment; DiD_f denotes the effect of receiving the first treatment f periods ago; and δ_f^D represents the average number of treatment the groups received for the first time f periods ago.

To ensure that the estimates can be representative for the treatment effect with equation (1), the following 5 assumptions must be satisfied:

Assumption 1: The treatment variable is at the group level, such as the treatment from a state law;

Assumption 2: No anticipation hypothesis that future treatments do not affect the current outcome;

Assumption 3: Parallel trend hypothesis that the treatment group and control group evolve with the parallel trend if there is no treatment;

Assumption 4: The treatment intensity for a specific group is either always weakly higher or always weakly lower than the intensity of its first treatment;

Assumption 5: At a period when at least one group increases/decreases its treatment intensity, there is at least another group with no change in its treatment intensity.

The estimates will be measured on a yearly basis at municipality level. The time and municipality fixed effects will be considered and the standard errors will be clustered at municipality level.

3.1 Treatment Variable

As illustrated in Section 2, municipalities have the autonomy to decide how many FHTs will participate. Therefore, the treatment is not binary in this setting. With the requirement for treatment to be discrete, the treatment intensity will be categorized into 3 groups based on the participation intensity of the municipalities: Intensity 1: 0-33.33%, Intensity 2: 33.34%-66.66%, and Intensity

3: 66.67%-100%. The participation intensity for a specific municipality is measured as the number of FHTs that participated in PMAQ divided by the total number of FHTs under its jurisdiction. The number of FHTs is the sum of the family health strategy teams and primary health teams.

3.2 Time-Varying covariates

The time-varying covariates include the primary care coverage, number of doctors, number of hospital beds, FHT quality ranking, total population, GDP per capita, education index, Bolsa familia coverage (i.e., BFC) and health transfer per capita.

One thing worth noting is that some of these variables, for example, the primary care coverage, number of doctors, number of hospital beds, FHT quality ranking and the health transfer per capita, can be affected by the PMAQ program. This may raise the concern for bad control problem. However, this study explores the effect of Monitoring and Evaluation, one of the interventions of the program, instead of the whole program. The intervention of only monitoring and evaluation is assumed to be orthogonal to the above-mentioned variables. The issue related to confounding effect from the other PMAQ interventions will be discussed in Section 3.2.2.

3.2.1 Self-Selection Bias

In the PMAQ setting, participation is voluntary and municipalities decide which team to register. In this situation, municipalities may select the FHTs with higher service quality to participate. Besides, the design of the instantaneous financial incentives would increase the likelihood that municipalities with financial constraints, especially in health construction, would register for the PMAQ program. In this circumstance, usually the matching methods (e.g., propensity score methods) will be employed. However, due to the big difference between the number of registered municipalities (treatment) and non-registered municipalities (control), especially for the second round (5072 vs 494) and the third round (5324 vs 242), the PSM will result in the sacrifice of a large amount of the sample. Therefore, the only approach used in this study to reduce the selection bias is to include the indicators related to the quality of the FHTs and the socioeconomic characteristics of the municipalities. The PMAQ score has been applied in most of the literature related to the PMAQ program as the quality measurement indicator (e.g., Kovacs et al., 2021; Fardousi et al., 2022), and it is publicly available from the Ministry of Health. However, during the period of this study, the official website[4] was inaccessible due to system maintenance, so I construct another quality indicator based on the amount of funds transferred per FHT for each municipality. This approach splits the municipalities into 5 ranking groups (1-5) with 1 indicating the municipalities with lowest-quality FHTs on average. Municipalities not enrolling in the PMAQ will be automatically assigned 1, based on the assumption that municipalities tend to avoid the participation of low-quality FHTs. For the socioeconomic and demographic characteristics, the total population, GDP per capita, education index (FIRJAN Index of Municipal Development for

[4] The website address is http://dab.saude.gov.br/portaldab/cidadao_pmaq2.php

education (i.e., IFDM)) and Bolsa familia coverage will be included. As the FIRJAN Index is only available for the years up to 2016, linear extrapolation has been used to generate the value for the years from 2017 to 2019.

3.2.2 Confounding Effects from Other PMAQ Interventions

Apart from the monitoring and evaluation, there are other interventions from the PMAQ program, for example, financial incentives, continuing education and institution support. The effect of financial incentives will be captured by the indicator health transfer per capita, since all the PMAQ funds must be spent on health. As discussed in Section 2, the continuing education and institution support are not compulsory and their progress will be assessed in the external evaluation process, which will be reflected in the PMAQ score. Therefore, the constructed indicator for FHTs' quality mentioned in the Section 3.2.1 will be employed to capture the effect from these two interventions.

3.2.3 Confounding Effect from Other Programs

During the implementation period of the PMAQ program, there were also other national health programs related to the PHC in effect. The Ministry of Health provides a detailed program list[5]. The ACSC hospitalization rate will be used in this study to construct the inequality index (the details can be found in Section 3.3.1). In the literature, there is evidence for the effect of PHC expansion on the ACSC hospitalization rate (Macinko et al., 2010) and on racial health inequality

5 Program list is available from https://www.gov.br/saude/pt-br/acesso-a-informacao/acoes-e-programas.

(Hone et al., 2017). Therefore, these programs may influence racial health inequality through their impact on PHC expansion.

There are also other channels for the confounding effect. For example, the More Doctors Program launched in 2013 was designed to increase the number of doctors in Brazil, which may affect the ACSC hospitalization rate by alleviating the shortage of human resources. To capture these effects, the primary care coverage, the number of doctors and the number of hospital beds are included.

3.3 Design for General and Specific Analysis

As mentioned in Section 2, monitoring indicators can be classified into seven categories, and four of them can be investigated through the indicators with the race-identification record in the database: women's health, children's health, diabetes mellitus and hypertension, and tuberculosis and leprosy. There is no suitable racial-identification indicator recorded for the other three modules. Therefore, the effect of monitoring and evaluation on racial health inequality will be investigated both generally at an overall level and specifically with the indicators related to those four categories. Moreover an age subgroup analysis will be added to the general analysis and the patients will be divided into three subgroups (0-4 years; 5-19 years; 20-64 years), following Russo et al.(2021).

Although researchers have reached a consensus on the health inequality between white and non-white people (e.g., Arias et al.,2004; Lovell, 2006; Bailey et al., 2013; Lima and Prates, 2018), the existence of an inequality between black and brown people has been a subject of debate during the last twenty years (e.g., Bailey et al., 2013). Considering the problem of self-reclassification of the race and the disparity in the difficulty regarding racial mobility between different racial groups (i.e., the transition from black to brown is easier to achieve than the transition from black/brown to white), this study will apply the dichotomous approach that the difference between only white and non-white individuals will be explored. The outcome variable will be an inequality index, while the indicators to calculate the index vary between general analysis and specific analysis. Since the comparison is based on two racial subgroups, a simple difference approach is preferred (Schlotheuber and Hosseinpoor, 2022).

3.3.1 General Analysis

In the general analysis, the ambulatory-care-sensitive conditions (i.e., ACSC) hospitalization rate will be used as the outcome variable. The ACSC hospitalization rate has been a popular indicator for the quality of PHC in the literature, as it is preventable and sensitive to the change in the quality of ambulatory care (Kim et al., 2019). One limitation of this indicator is that it will be difficult to distinguish whether the effect originates in a change in the health condition of the group it measures or from an improvement in the healthcare service.

The ACSC hospitalization rate is adjusted for age and race with the direct method of standardization (Pan American Health Organization, 2018), which is measured as the ACSC hospitalization of patients per 1000 for one specific age group and one specific racial group. The inequality index calculated with the ACSC hospitalization rate can represent two aspects of racial health inequality: firstly, it indicates the difference in the quality of PHC received by two different racial groups; secondly, it can serve as an indicator for the disparity in the health condition itself. A higher ACSC hospitalization rate indicates lower PHC quality and poorer health. The inequality index is constructed with the following equation:

$$Inequality_{i,t} = NWACSC_{i,t} - WACSC_{i,t} \qquad (2)$$

Where $Inequality_{i,t}$ denotes the racial health inequality for municipality *i* in year *t*; $NWACSC_{i,t}$ represents the age-race-adjusted ACSC hospitalization rate for the non-white group for municipality *i* in year *t*, and $WACSC_{i,t}$ is the age-race-adjusted ACSC hospitalization rate for the white group for municipality *i* in year *t*. It is because the white group is assumed to have better health and receive a higher quality of PHC than the non-white group.

3.3.2 Specific Analysis

In the specific analysis, Three other different indicators will be applied: 1) For children's health, the inequality index is constructed with the ACSC hospitalization rate in the same way as equation (2), however, only the outcome of 0-4 years will be explored; 2) For women's health (especially pregnant women), the average number of pre-natal consultations is used and the inequality index is measured as the outcome for the white group minus the outcome for the non-white group[6], and 3) The index of diabetes mellitus and hypertension/tuberculosis and leprosy (i.e., DHTL in the figures and tables) is combined into one indicator, which is the hospitalization rate for these four types of illnesses. Since these four illness categories are also included in the ACSC causes, they can reflect the change in PHC quality and health condition. The ACSC is classified according to the ICD-10 and the details of the illnesses included in ACSC are available from the Brazilian Ministry of Health[7].

4. Sample Selection

This study explores the racial health inequality in Brazil at municipality level from 2009 to 2019 on a yearly basis. Following Russo et al. (2021), the sample exclude individuals aged higher than 64 years old, since they are more likely to be affected by other diseases, which may not be healed by timely and high-quality PHC. The municipalities in the treatment group with both higher and

[6] Different from the ACSC hospitalization rate that lower hospitalization implies better health condition and higher quality of treatment and service, for the pre-natal consultations, higher number indicates better health outcome. Therefore, white group is assumed to have higher number of average pre-natal consultations than non-white group.

[7] ACSC regarding the ICD-10 is available from https://bvsms.saude.gov.br/bvs/saudelegis/sas/2008/prt0221_17_04_2008.html

lower treatment intensity than their first switch intensity in the later rounds are excluded to satisfy the Assumption 4. Moreover, I also exclude the municipalities that drop out of the PMAQ programs to ensure that the parallel trend hypothesis is not violated. It is worth noting that not all patients have their race recorded in the SIH system. For the robustness of the estimates, the sample will only be included if the percentage of patients with missing race information is less than 10% of the total hospitalization number.

5. Data source

The data in this study comes from different databases. The ACSC hospitalization rate is available from the Hospital Information System (SIH/DATASUS). The ACSC mortality rate is extracted from the Mortality Information System (SIM/DATASUS). The data and information related to the PMAQ program, the number of FHTs and the primary care coverage can be found from the Ministry of Health. The number of pre-natal consultations is taken from the Brazil Live Birth Information System (SINASC/Ministry of Health). The BFC is available from the PBF Health System (BFA/DATASUS) before 2018 and (BFA/Ministry of Health) after 2018. The number of doctors and hospital beds are available from the CNES/DATASUS. The health transfer per capita can be found in the Information System on Public Health Budgets (SIOPS/DATASUS). The FIRJAN Index is published on IFDM official website. GDP per capita and total population are available from IBGE while the age-adjusted population is from preliminary estimates prepared by the Ministry of Health (SVS/DATATUS). The race ratio for each age subgroup is calculated with data from the 2010 Brazil census. The descriptive data for baseline (i.e., 2011) is listed in Table 2 as follows:

Table 2 Descriptive Statistics for Baseline Year (2011)

	Obs.	**Mean**	**Stan Dev**	**Min**	**Max**
Outcome Variable					
General Analysis (ACSC Hospitalization Rate)					
0-64 years					
Inequality	2091	-0.75	15.6	-61.51	118.12
White	2091	10.95	10.07	0	77.58
Non-White	2091	10.20	11.55	0	119.06
	Obs.	**Mean**	**Stan Dev**	**Min**	**Max**
0-4 years (Children's Health)					
Inequality	2263	-1.31	49.59	-208.15	530.85
White	2263	28.77	30.92	0	239.62
Non-White	2263	27.46	38.94	0	552.08
5-19 years					
Inequality	2263	-0.37	10.24	-72.10	90.98
White	2263	5.47	7	0	86.38
Non-White	2263	5.10	7.75	0	90.98
20-64 years					
Inequality	2112	-0.31	15.91	-82.59	120.28
White	2112	10.74	10.45	0	90.31
Non-White	2112	10.43	11.83	0	120.67
Specific Analysis					
Children's Health (0-4 ACSC Hospitalization rate)					
Inequality	2263	-1.31	49.59	-208.15	530.85
White	2263	28.77	30.92	0	239.62
Non-White	2263	27.46	38.94	0	552.08
Women's Health (Number of Pre-natal Consultation)					
Inequality	4953	0.089	0.29	-3.125	3.07
White	4953	3.59	0.33	1	6.50
Non-White	4953	3.50	0.34	1	6.50
DHTL (DHTL Hospitalization rate)					
Inequality	2238	0.11	2.09	-14.08	12.80
White	2238	1.14	1.45	0	14.79
Non-White	2238	1.25	1.71	0	18.78

Time-Varying Covariates					
Primary Care Coverage	2091	0.86	0.24	0	1
Total Population	2091	23982.16	70001.81	824	1900000
Number of Hospital Bed	2091	58.42	228.05	0	7381.42
Number of Doctor	2091	30.20	181.79	0	6046.67
GDP Per Capita	2091	15897.91	14428	2794.76	194926.10
Health Transfer Per Capita	2091	446.79	198.92	0	2045.27
Education Index	2091	0.75	0.12	0.28	1
Bolsa familia coverage	2091	0.79	0.15	0	1
FHT Quality Rank	2091	2.31	1.49	1	5

The descriptive data is collected for baseline years (all assumed to be 2011). The inequality for all (except women's health) is constructed with the difference between the ACSC rate for the related illness for non-white and white. The inequality for women's health is constructed with the difference of the number of pre-natal consultation between white and non-white individuals.

6. Result

6.1 General Analysis

6.1.1 General Analysis Based on Full Sample 0-64 years

The investigation into the effect of monitoring and evaluation starts with the general analysis based on full sample (0-64 years) and the results are exhibited in the Figure 1. Unlike the positive trends in the pre-treatment period, the monitoring and evaluation started to affect inequality with an immediate but small negative effect (a decrease of 0.35 ACSC hospitalization rate in the difference between white and non-white with one more unit of treatment intensity) and the magnitude of the effect kept increasing to about 2.5 until the third year; it remained constant in the following periods. To explore the cause, I test how the treatment affects white and non-white subgroups and show the results in Figure 2. Different from the expectation that monitoring and evaluation could contribute to lower the ACSC hospitalization rate, I find that the ASCS hospitalization rate for white shifted up with higher treatment intensity and the magnitude rose continuously from 0.3 to 1.8 during the 5-year period. For non-white individuals, the increase in treatment intensity appeared to reduce the ACSC hospitalization rate from the first year of post-treatment period and the effect fluctuated between 0.65 to 1.72. Therefore, the reduction in inequality is driven by both the growth in the ACSC hospitalization rate among white people and the reduction in the ACSC hospitalization rate among non-white people.

Figure 1 General Analysis Trend Based on Full Sample (0-64 Years)

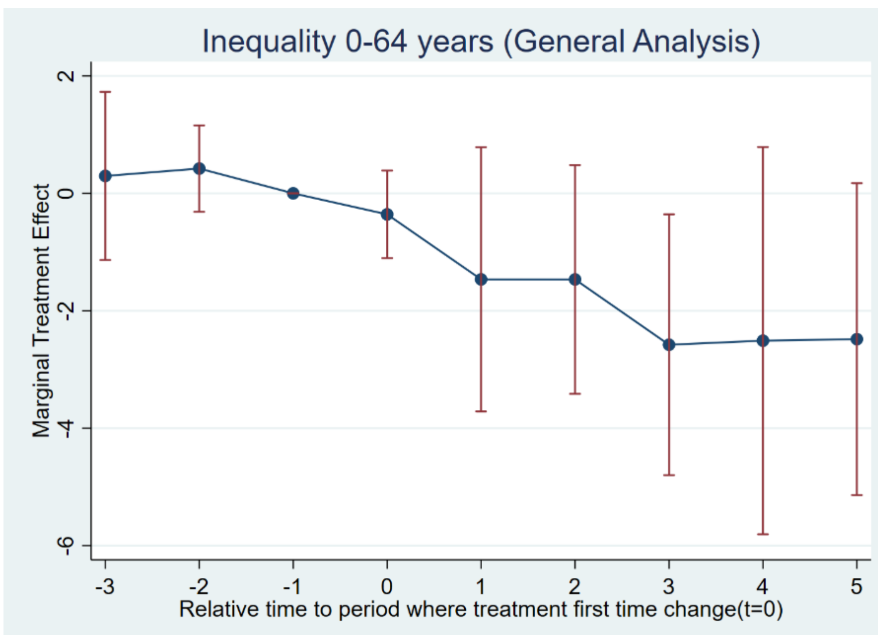


Figure 1 displays the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 0-64 years sample for inequality constructed with ACSC hospitalization rate. The standard errors are clustered at municipality level.

Figure 2 Racial Subgroup Trend Based on Full Sample (0-64 Years)

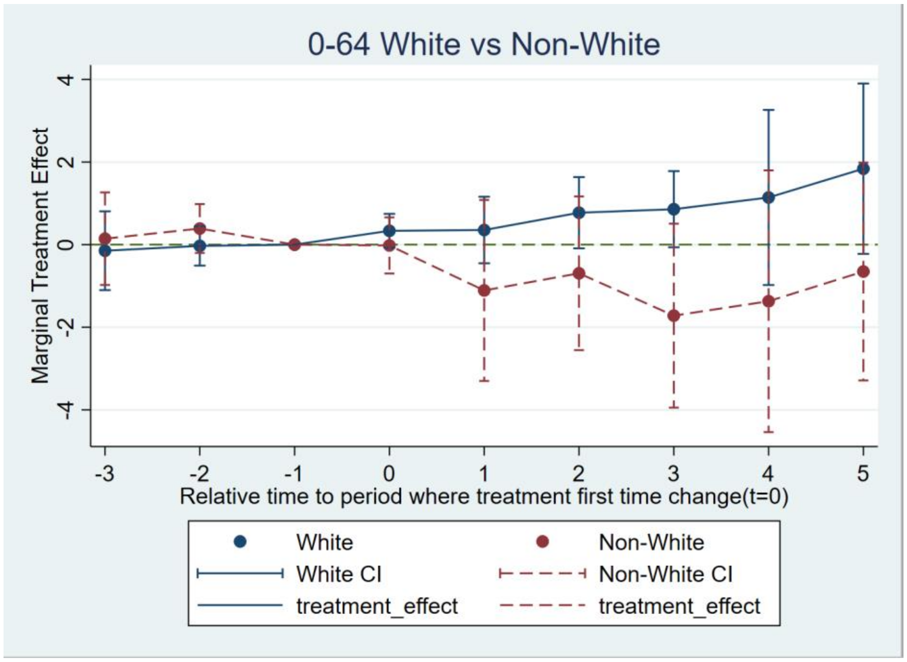


Figure 2 presents the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 0-64 years sample for ACSC hospitalization rate of white (Blue) and non-white (Red) groups. The standard errors are clustered at municipality level.

6.1.2 General Analysis Based on Age Subgroups

In addition to the general analysis based on the full sample, I conduct a subgroup analysis to inspect the effect of treatment for different age groups and divide the sample into three age-based subgroups (0-4 years, 5-19 years, 20-64 years) following Russo et al. (2021). As shown in Figure 3, the marginal treatment effect on racial health inequality is negative for all three groups. The 0-4 years subgroup experienced a boom in the magnitude of the marginal effect during the third year from 0.6 to 8.1 and converged to the same level as the 20-64 years subgroup thereafter, while the marginal effect on the 5-19 years subgroup appeared to be the gentlest. Due to the significant difference in the confidence interval of the estimates between 0-4 years and the other two age subgroups, and the function of the 0-4 years ASCS hospitalization rate for children health investigation, the results for 0-4 will be separately displayed and analyzed in Section 6.2 for the specific analysis. The results for 5-19 and 20-64 are combined and explored here (see Figure 4). For 5-19 years, the marginal treatment effect remained small in the short-run (less than 0.41) until year 3. It began to increase from 0.8 to 2.1 in year 4, followed by a slight drop to 1.4 in year 5. The 20-64 years experienced an instantaneous negative marginal effect of 1.1, which kept growing in the following periods to 3.9 in year 5. When further exploring the racial subgroups (5-19: Figure 5; 20-64: Figure 6), we can see that the decrease in inequality for the population aged 5-19 is mainly the result of the positive marginal treatment effect on ACSC hospitalization rate for white individuals, which elevated from year 2 and fluctuated between 0.7 to 1.7 afterwards, while the ACSC hospitalization rate for non-white individuals witnessed little marginal treatment effect. As for the racial groups aged 20-64 years, the marginal treatment effect rose immediately from the first year and expanded during most post-treatment periods for both white (positive) and non-white (negative). The magnitude for both reached around 2 after receiving treatment for 4-5 years. The estimates and the standard errors for the general analysis and the age subgroups are listed in Appendix 1 and Appendix 2 respectively. We can observe from Appendix 1 that the differences between the estimates with and without controls for most post-treatment periods are less than 10%, which implies the small effect from the unobservables (Altonji et al., 2005).

Figure 3 Age Subgroups for General Analysis (0-64 Years)

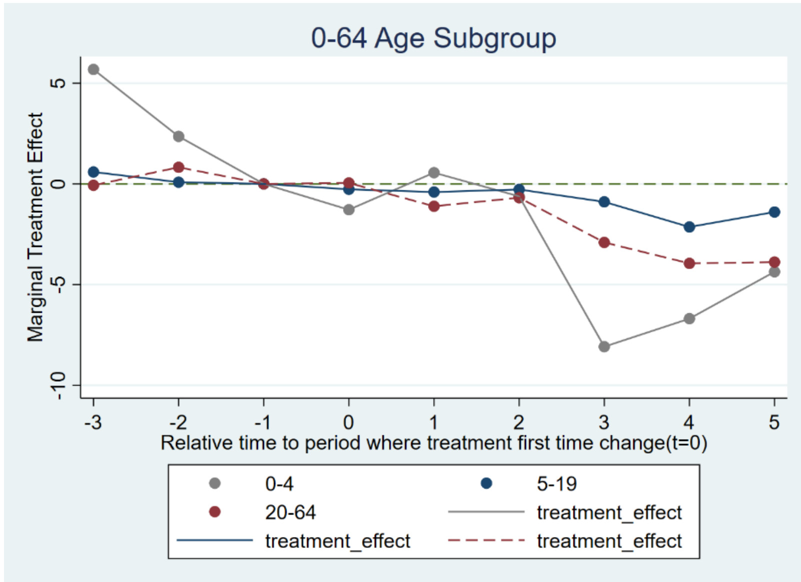


Figure 3 presents the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates for inequality constructed with ACSC hospitalization rate based on 3 age subgroups: 0-4 (Gray), 5-19 (Blue), 20-64 (Red) without confidence interval. The standard errors are clustered at municipality level.

Figure 4 Age Subgroups for General Analysis (5-64 Years)

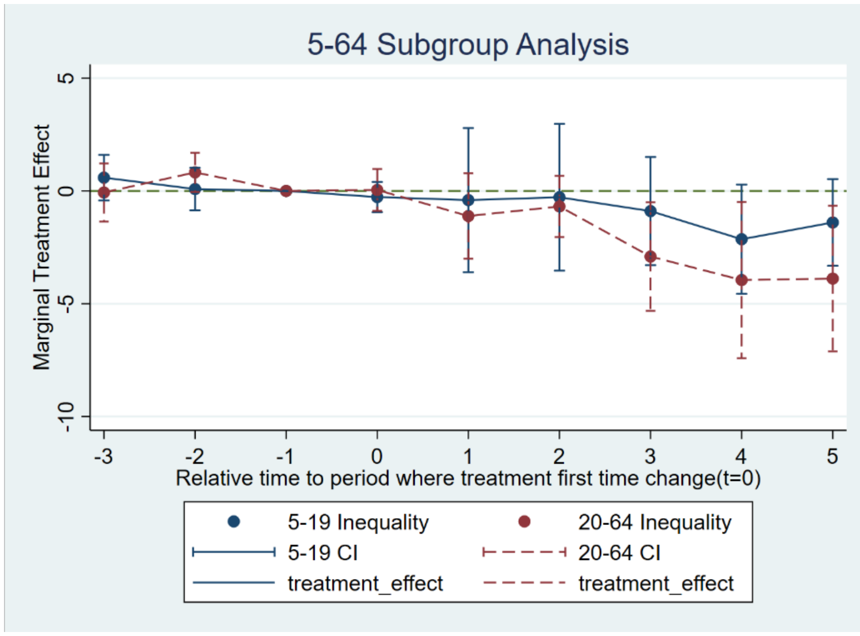


Figure 4 presents the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals for two age subgroups: 5-19 (Blue), 20-64 (Red). The standard errors are clustered at municipality level.

Figure 5 Racial Subgroup Trend Based on 5-19 Years

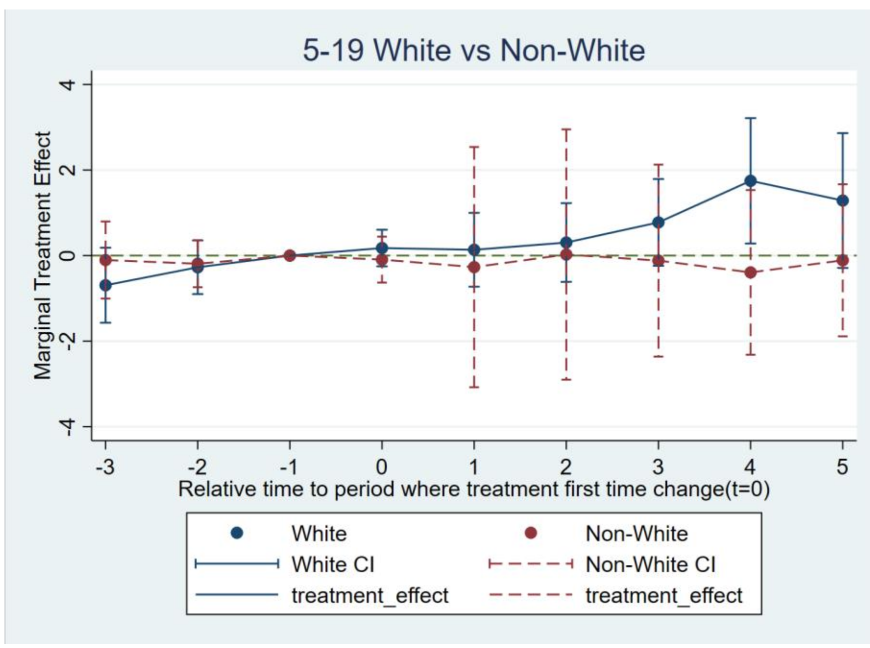


Figure 5 presents the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 5-19 subgroup for ACSC hospitalization rate of white (Blue) and non-white (Red). The standard errors are clustered at municipality level.

Figure 6 Racial Subgroup Trend Based on 20-64 Years

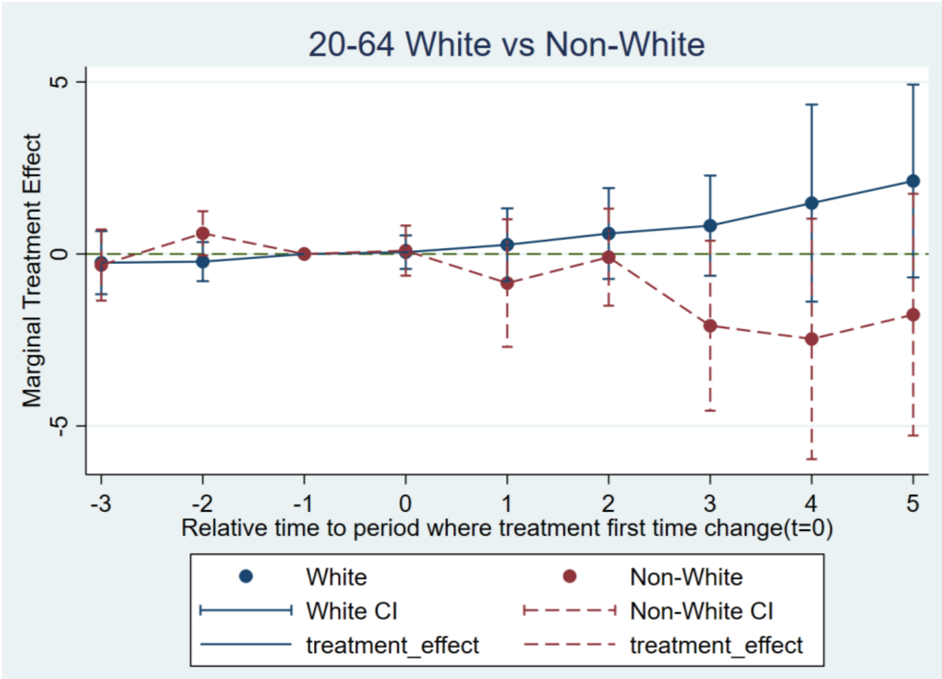


Figure 6 presents the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 20-64 subgroup for ACSC hospitalization rate of white (Blue) and non-white (Red). The standard errors are clustered at municipality level.

6.2 Specific Analysis

To further examine the effect of monitoring and evaluation, I investigated other three indicators related to the four modules that were assessed during the monitoring and external evaluation process: children's health, women's health, diabetes mellitus and hypertension, and tuberculosis and leprosy.

6.2.1 Children's Health (0-4 Years)

I used the ACSC hospitalization rate for 0-4 years to construct the racial health inequality index for children's health and the results are presented in Figure 7. There was little response to the treatment for the first two periods, followed by a sharp increase to 8.1 in year 3, and a decline in the last two periods to 4.4 in year 5. Increasing the treatment intensity benefited both white and non-white during the first 3 years while the marginal treatment effect on ACSC hospitalization for white individuals reversed the direction to become positive in year 4 and increased to 5.7 in year 5 (see Figure 8). The effect for the non-white group reached a peak in year 3 (8.1) and kept dropping afterwards. The treatment effect even changed the direction to be positive (1.3) in year 5. The trend may be caused by variance in attitudes towards children of different races in different municipalities, especially for non-white children in the later periods. This inference could be bolstered by the results that there was the large confidence interval (-26.4 to 19.6 for non-white in year 4) for the marginal treatment effect, while the standard errors for white children were less than half of that for their non-white peers (4.3 for white and 11.4 for non-white). The estimates and standard errors for children's health (0-4) are presented in Appendix 2 together with another 2 age subgroups.

Figure 7 Specific Analysis Trend Based on 0-4 Years: Children's Health

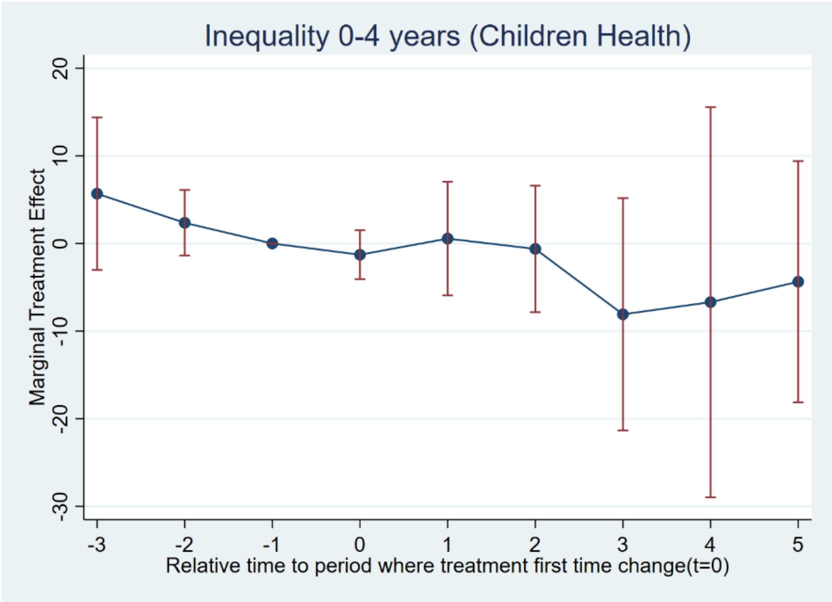


Figure 7 displays the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 0-4 years subgroup for inequality constructed with ACSC hospitalization rate. The standard errors are clustered at municipality level.

Figure 8 Racial Subgroup Trend Based on 0-4 Years: Children's Health

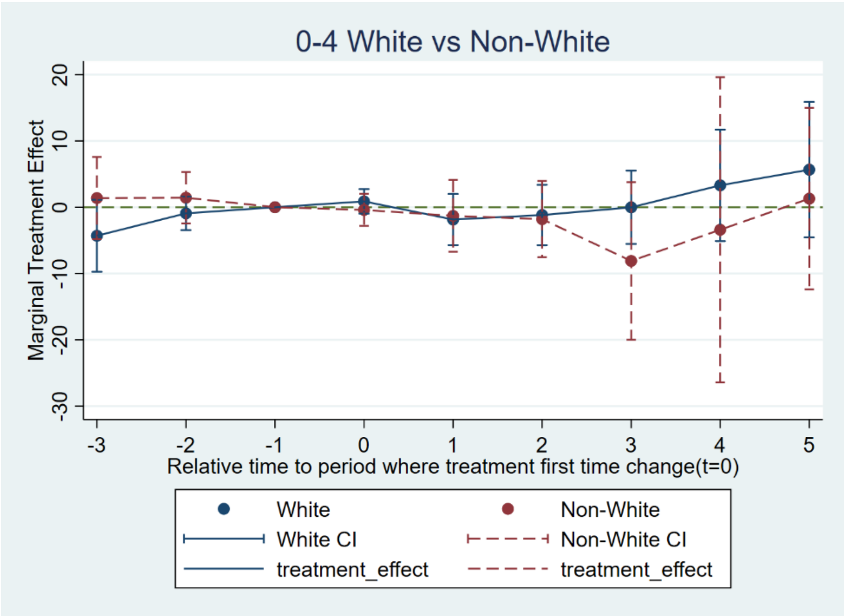


Figure 8 presents the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 0-4 subgroup for ACSC hospitalization rate of white (Blue) and non-white (Red). The standard errors are clustered at municipality level.

6.2.2 Women's Health

Pregnant women are a key demographic for women's health monitoring, so here I employ the number of pre-natal consultations to calculate the inequality index and the higher number of consultations indicates healthier status. Figure 9 shows that there was an instantaneous negative marginal treatment effect on inequality (0.016 in year 0), which remained relatively constant between 0.02 and 0.235 in the later years. As for the two racial groups (Figure 10), the non-white group experienced an immediate rise in the number of pre-natal consultations and the effect presented an upward trend afterwards (from 0.09 in year 0 to 0.4 in year 5). The marginal effect for white people suffered a slight drop immediately (-0.0068 in year 0) and the treatment displayed little effect until year 3, followed by a shift to the positive marginal effect of 0.028 afterwards. The racial inequality reduction is mainly caused by the improvement in non-white pre-natal consultations, and white women also began to benefit after around 3 years.

Figure 9 Specific Analysis Trend: Women's Health

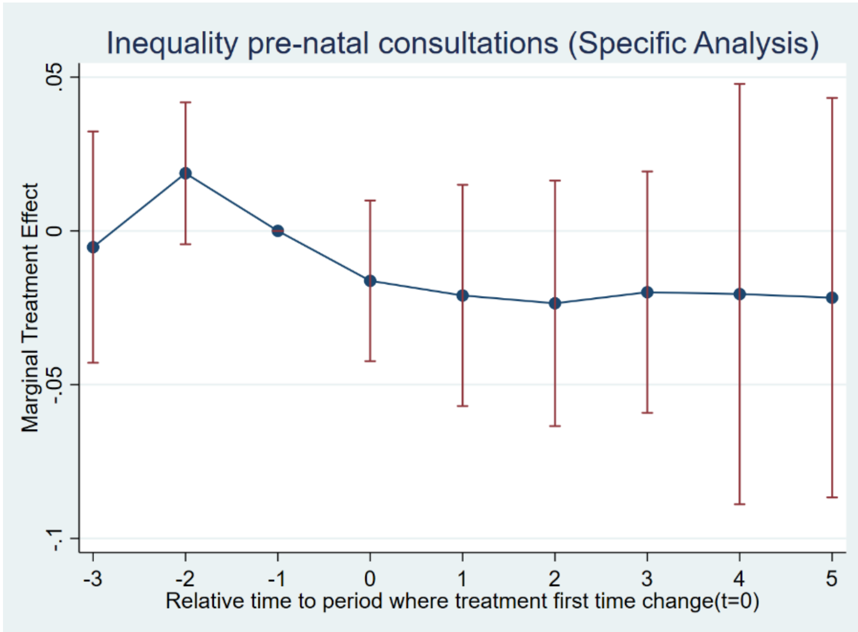


Figure 9 displays the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals for inequality constructed with the number of pre-natal consultation. The standard errors are clustered at municipality level.

Figure 10 Racial Subgroup Trend: Women's Health

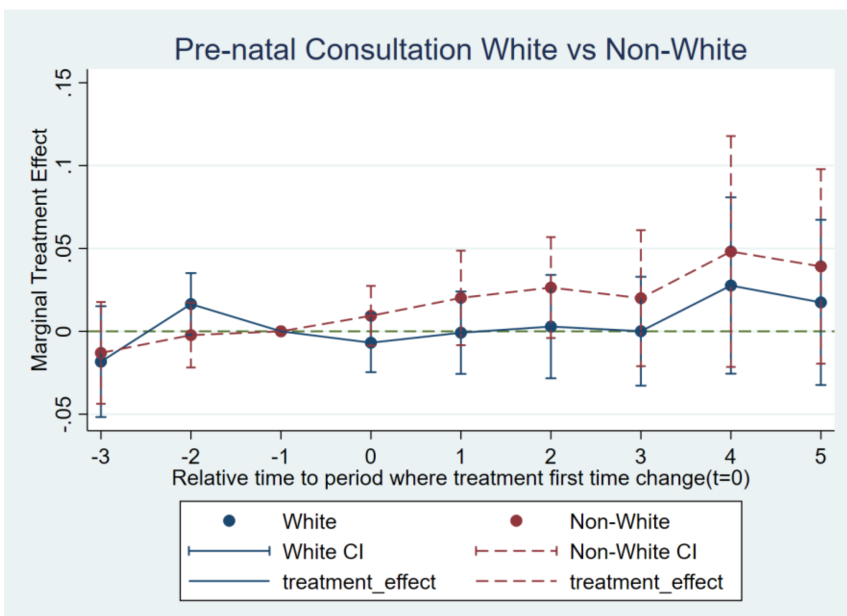


Figure 10 presents the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals for average pre-natal consultation number of white (Blue) and non-white (Red). The standard errors are clustered at municipality level.

6.2.3 Diabetes Mellitus and Hypertension, and Tuberculosis and Leprosy (0-64 Years)

The Diabetes mellitus and hypertension, and Tuberculosis and leprosy (i.e., DHTL) are two other important modules for monitoring and external evaluation. They are both included in the ACSC list, so I use the hospitalization rate for these four illness categories to measure the inequality index. The results are displayed in Figure 11. There was an instant negative marginal effect (0.099 in year 0) in inequality related to the DHTL causes, and an upward trend in magnitude to 0.616 in year 5. When it comes to the two racial groups (Figure 12), the marginal treatment effect for white people remained slight until year 3 (when it increased to 0.177) and stayed relatively constant afterwards, while non-white people experienced an immediate negative marginal treatment effect, which expanded in the following years to 0.45.

The estimates and standard errors for women's health (pre-natal consultation) and diabetes mellitus and hypertension, and tuberculosis and leprosy (DHTL) are presented in Appendix 3.

Figure 11 Specific Analysis Trend Based on Full Sample (0-64 Years): DHTL

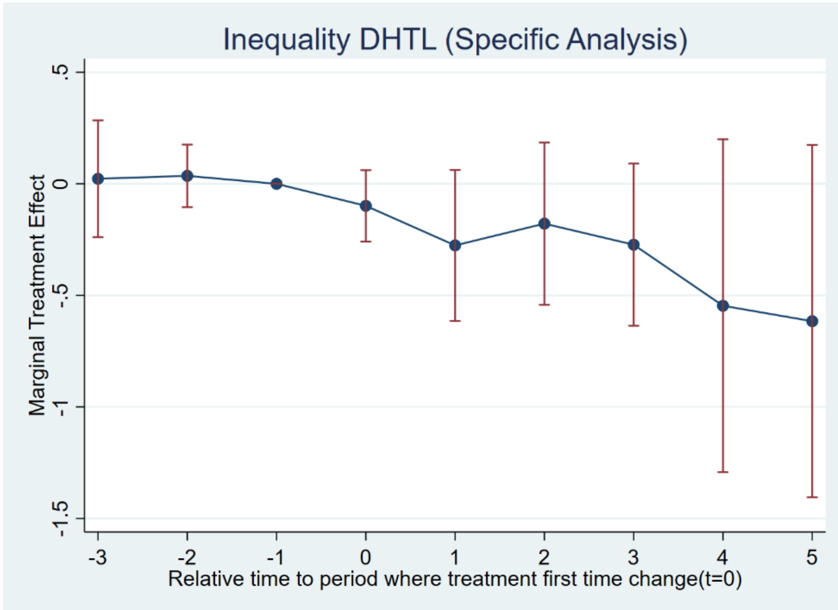


Figure 11 displays the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 0-64 years sample for inequality constructed with the hospitalization rate of Diabetes Mellitus, Hypertension, Tuberculosis and Leprosy. The standard errors are clustered at municipality level.

Figure 12 Racial Subgroup Trend Based on Full Sample (0-64 Years): DHTL

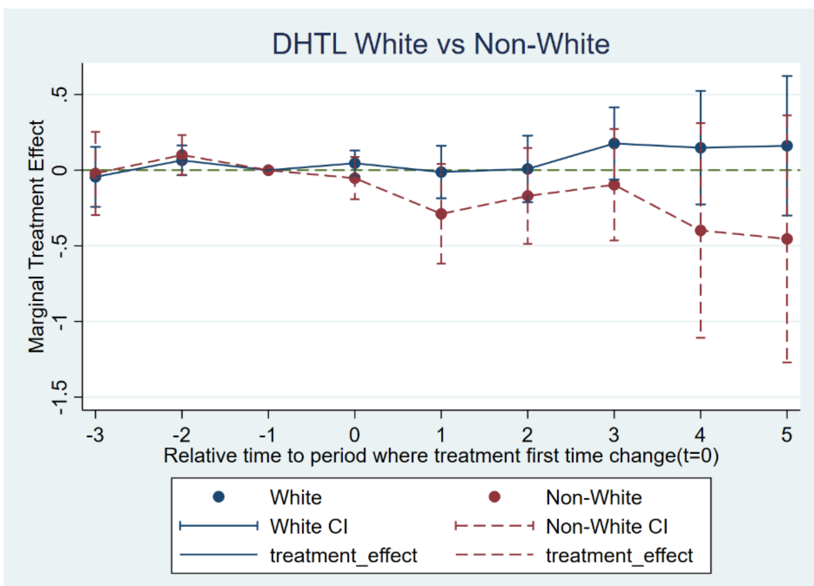


Figure 12 presents the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 0-64 for DHTL hospitalization rate of white (Blue) and non-white (Red). The standard errors are clustered at municipality level.

6.3 Discussion of Results

Overall, the results support the impact of monitoring and evaluation on reducing racial health inequality. This finding is bolstered by the age subgroup analysis that the subgroup with specific monitoring requirement (0-4) contributed more to the marginal treatment effect than the other two age subgroups, especially after the second treatment year (see Figure 3).

However, it is worth noting that increasing the intensity of the monitoring and evaluation raised the ACSC hospitalization rate for white individuals, which is opposite to what is expected to happen with the higher PHC service quality brought about by the PMAQ program. This counterintuitive positive marginal effect may arise from two channels: Firstly, it may be caused by the decrease in racial discrimination in the health system, and the extra effort to improve the health condition of non-white individuals. As indicated in the literature, non-white people tend to use the health service much less than white people do (e.g., Paixão, 2010), and this may be partially due to the racial discrimination in health resource allocation. When there is less racial discrimination in the health system, there will be a more equal distribution of resources away from white people towards non-white people, especially when resources are limited. Therefore, in the transition, white people's access to health resources would be reduced, especially when the extra benefits brought by the higher quality and efficiency of PHC are not enough to reduce racial discrimination and improve healthcare for both racial groups at the same time. When the force to reduce the racial discrimination is strong, this change would be reflected in the upward trend of the ACSC hospitalization rate for white people when the treatment intensity increases. If this channel dominates, then the results support the effect of monitoring and evaluation on reducing the racial health inequality.

The second potential channel is the racial discrimination in the health system. The record of ACSC hospitalization depends on resource allocation in the hospital and if there is racial discrimination, white people will have priority in hospitalization admission approval. As a result, more hospitalization applications from white patients will be approved, with a higher turn-over rate of the hospital beds brought about by the higher PHC quality and higher efficiency in service delivery, after the monitoring and evaluation intervention. If this channel contributes to most of the effect, then benefits from the treatment would favor white people, and the intervention to improve PHC quality would increase racial health inequality because more non-white people would have their hospitalization applications rejected due to the racial discrimination.

I conduct a robustness check to test this issue (see Section 7.2), which eliminates the concerns that the racial discrimination may dominate the effect. In addition, Chaisemartin and D'Haultfoeuille (2022) set the baseline year to be time=-1 for the placebo test, and the parallel trend assumption can be examined by observing the trend during the pre-treatment periods. The results for analysis of all the groups pass the placebo tests at the 95% significance level.

7. Robustness Check

7.1 Misclassification in Race

As described in the introduction, racial identifications are different in institutional records and self-classification in Brazil. Racial classification in institutions depends on skin color, while self-reported race in census is based on both socioeconomic status and self-identification. This can result in a mismatch of race records between the Hospital Information System (SIH) and self-reporting in the census. From the literature, the likelihood of self-reported white patients being registered as non-white in SIH is much higher than the opposite outcome (e.g., Carvalho et al., 2004; Schwartzman, 2007; Hone et al., 2021). This mismatch will introduce bias to the estimates via a measurement error in the ACSC hospitalization rate. However, the extent of the misclassification error is not random among the municipalities. As described in the introduction, the declining social mobility may serve as an obstacle for the transition between race. Therefore, I measure social mobility based on the growth of GDP per capita from 2002 to 2010 and divide the municipalities into three groups, assuming that the group with higher growth rate experienced higher social mobility, and thus should be exposed more to the misclassification error. I re-assign 5%, 10% and 15% of non-white patients to be white patients respectively, according to their growth rate group, and re-estimate the full sample (0-64 years) and the racial subgroup. The results and estimates are listed in Appendix 4, Appendix 5 and Appendix 6. The results show similar trends to the general analysis (0-64 years) regarding both inequality and the two racial subgroups, and thus strengthen the robustness of the finding.

7.2 Racial Discrimination in Health System

From the discussion in Section 6.3, racial discrimination may also be a path for the observed "plausible" inequality reduction. To test if this path dominates the treatment effect, I replace the ACSC mortality rate with the ACSC hospitalization rate as the indicator to construct the inequality index. To ensure the robustness of the result, only the sample with less than 6% of missing information for race is included. The reason to use the ACSC mortality rates is that the record of the mortality rate is not restricted by the availability of resources in the institutions or the racial discrimination in resource allocation, which can help disengage the effect. If the racial discrimination in the hospital dominates and the reduction in ACSC hospitalization for non-white individuals is caused by the biased resource allocation, then the health status for non-white individuals and the quality of the PHC they received should be lower than their white peers. In this case, the racial health inequality constructed with the ACSC mortality rate would experience a positive marginal treatment effect. If the reduction in the racial discrimination dominates, then the marginal treatment effect on inequality with ACSC mortality rates should be negative. However, due to the fact that mortality rate is an extreme indicator, it is assumed to be not sensitive to the treatment. Therefore, the change in the latter years would be more likely to reflect the marginal treatment effect. Besides, better race reporting in the Mortality Information System creates a big enough sample to extend the post-treatment period to year 6. The results are listed in Appendix 7. We can observe from the results that the marginal treatment effect on racial inequality stayed close to zero during the first 4 years and started to increase in magnitude to 0.077 in the following 2 years, which is consistent with the insensitiveness of the mortality rate. This result suggests that the reduction in racial discrimination is the main cause for the findings with the ACSC hospitalization rates, which can be further strengthened by the racial subgroup trends for the ACSC mortality rate from 2009 to 2019 (see Appendix 8). Though the upward trend continued for non-white people after the start of round 1 of PMAQ, the downward trend for white people did not fall below 0.46 after 2014, followed by an overall increase afterwards, which was in contrast to their trend in the pre-treatment period. This trend indicates that the growth in health status and PHC services for white people was overall no higher than in their pre-treatment period and rules out the possibility that racial discrimination in resource allocation dominates the effect. The estimates and standard errors are displayed in Appendix 6.

7.3 No Anticipation Hypothesis

As shown in Section 3, one significant requirement for the robustness of the estimates is the No Anticipation Hypothesis, that the current outcome will not be affected by the anticipation of future treatments. Therefore, a placebo test provided by de Chaisemartin and D'Haultfoeuille (2022), which examines whether the parallel trends hold for 2 consecutive periods, is applied to test the no anticipation hypothesis. If the anticipation effect exists, then the parallel trend for 2 consecutive periods will be violated and the placebo estimates will show a significant deviation from 0. The results for all group analysis are exhibited in Appendix 9 and there is no placebo test coefficient significantly different from 0 at 5% significance level. This meets the no anticipation requirement for the robustness of the estimates.

7.4 Heterogeneous Treatment Effect

One important reason for this study to abandon the simple difference-in-indifferences or event study approach and opt for de Chaisemartin and D'Haultfoeuille (2022) is the possibility of the existence of a heterogeneous treatment effect. Therefore, the sign of weight for the TWFE estimates is measured for both general and specific analysis and the results are displayed in Appendix 10. From the summary of the results we can see that a large number of negative weights exist for every group analysis, which backs up the existence of the heterogeneous treatment effect and the switch towards heterogeneity-robust estimators.

8. Conclusion

8.1 Summary

In this study, I explore how monitoring and evaluation affects racial health inequality based on the Program for Improving Access and Quality of Primary Care (PMAQ) at municipality level from 2009 to 2019 in Brazil. The impact of monitoring and evaluation is investigated both generally with all ACSCs causes and specifically with 4 strategic areas in the monitoring and evaluation process: children's health, women's health, diabetes mellitus and hypertension, and tuberculosis and leprosy. I also conduct an age subgroup analysis in addition to the general analysis. Instead of the classical difference-in-differences or event study, I employ the heterogeneity-robust TWFE estimator proposed by de Chaisemartin and D'Haultfoeuille (2022) to consider the heterogeneous treatment effect. The results suggest that increasing the intensity of monitoring and evaluation can reduce racial health inequality, and that the 0-4 age group accounts for the largest part of the marginal treatment, while the effect on 5-19 years is the smallest. Moreover, the quality of the PHC received by the non-white group, and their overall health also benefit from the monitoring and evaluation intervention. One interesting finding is that the increase in monitoring and evaluation has led to a higher ACSC hospitalization for white people, lasting at least 5 years after the treatment. It may be mainly driven by the extra effort to improve health conditions for non-white people, given that resources to improve health conditions for both white and non-white people remain limited. This inference can also be partially supported by the robustness check. Therefore, one policy recommendation is to increase the allocation of PHC resources (e.g., the number of PHC teams and number of community health agents) and to encourage efficient use of equipment to cooperate with the monitoring and evaluation intervention. This could ensure that non-white and white people can both benefit from treatment and racial health inequality can be reduced at the same time.

This study contributes to mainly two streams of literature. Firstly, it enriches the literature related to racial health inequality and provides novel evidence on how monitoring and evaluation affects racial health inequality in Brazil. Secondly, it expands the research on the effect of the PMAQ program and P4P strategies in Brazil.

8.2 Limitations

There are also some limitations in this study. The first limitation is the selection bias. Although some observables have been considered by including the control variables (Section 3.2.1), there may be other unobservables that influence municipalities' decisions regarding PMAQ participation, which may introduce bias to the estimates. One way to solve this problem is to interpret the estimates as ATE instead of ATT (Callaway et al., 2021). Secondly, the treatment intensity can only be explored with discrete instead of continuous metrics due to the limited options between discrete treatment intensity with non-staggered treatment and continuous treatment intensity with staggered treatment provided in de Chaisemartin and D'Haultfoeuille (2022).

However, there are no alternative heterogeneity-robust TWFE estimators with STATA packages available that can be used for the non-staggered, continuous treatment, time-varying covariates and dynamic effect. Thirdly, the representativeness of the results in this study is restricted to the municipalities with relatively complete race information in SIH (missing race information less than 10%) and the effect of the monitoring and evaluation treatment may vary between the municipalities with higher race report rates and lower race report rates. For example, the physicians in the municipalities with lower race report rates may not endeavor to stop the racial discrimination behavior even after they are monitored, as there is a high likelihood that the race of the patients may not be recorded. The third limitation lies in the measurement error in the ACSC hospitalization rate. The proportion of people in each race and age group from the 2010 census is used to calculate the ACSC hospitalization rate for the whole period, from 2009 to 2019, since the proportion of the population grouped by race and age at municipality level in Brazil is only available from the census. This measurement approach will result in a measurement error in the ACSC hospitalization rate, and the accuracy of the race ratio also decreases over time, which will introduce bias to the estimates. This bias can be partially eliminated after the publication of the 2022 Brazil census. The linear interpolation and extrapolation approach can be employed to estimate the population grouped by age and race for each year, using the 2010 and 2022 census. Another limitation is derived from the measurement error in FHTs' quality. The absolute PMAQ score is a direct indicator for the FHTs' quality, which should be publicly available from the Ministry of Health. However, due to system maintenance during this study period, the PMAQ scores are not available. Therefore, the amount of financial funds has been applied to construct the ranking. However, this indicator may not reveal the real FHT quality ranking since for the first two rounds, the amount of financial funds also depends on the socioeconomic status of the municipalities. This problem may also introduce bias to the result, since it is related to the selection bias issue. It can be resolved when the PMAQ scores become available.

There are several ways for future research to make improvements. Firstly non-parametric estimators can be employed to explore this relationship between monitoring and evaluation, and health outcomes. Secondly, there are some other factors, for example, bonus incentives for FHT staff, that may have a mediate impact on the effect of monitoring and evaluation. It is worth exploring this mediate impact using the related indicator in the PMAQ manager survey and the survey should be accessible when the system maintenance is over.

Bibliography

Alesina, A. and Rodrik, D., 1994. Distributive politics and economic growth. *Quaterly Journal of Economics*, 109(2): 465-490.

Alesina, A., Stantcheva, S., and Teso, E., 2018. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2): 521-554.

Altonji, J.G., Elder, T.E. and Taber, C.R., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1):151–184.

Arias, O., Yamada, G. and Tejerina, L., 2004. Education, family background and racial earnings inequality in Brazil. *International Journal of Manpower*, 25(3/4): 355-374.

Athey, S. and Imbens, G.W., 2022. Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 2261(1): 62–79.

Bailey, S., Loveman, M. and Muniz, J., 2013. Measures of "Race" and the analysis of racial inequality in Brazil. *Social Science Research*, 42(1):106-119.

Borusyak, K., Jaravel, X. and Spiess, J., 2021, Revisiting event study designs: Robust andefficient estimation. arXiv preprint arXiv:2108.12419.

Buetow, S., 2008. Pay-for-performance in New Zealand primary health care. *Journal of Health Organization and Management*, 22(1): 36-47.

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

Callaway, B., Goodman-Bacon, A. and Sant'Anna, P.H., 2021. Difference-in-differences with a continuous treatment. arXiv preprint arXiv:2107.02637.

Carvalho, J.A.M., Wood, C.H. and Andrade, F.C.D., 2004. Estimating the stability of census-based racial/ethnic classifications: the case of Brazil. *Population Studies* 58 (3): 331–343.

Castro, M., Massuda, A., Almeida, G., Menezes-Filho, N., Andrade, M., de Souza Noronha, K., Rocha, R., Macinko, J., Hone, T., Tasca, R., Giovanella, L., Malik, A., Werneck, H., Fachini, L. and Atun, R., 2019. Brazil's unified health system: the first 30 years and prospects for the future. *The Lancet*, 394(10195): 345-356.

Chor, D., 2013. Health inequalities in Brazil: race matters. *Cadernos de Saúde Pública.*

Covin, D., 2006. *The Unified Black Movement in Brazil*. Jefferson: MacFarland and Company.

de Chaisemartin, C. and D'Haultfœuille, X. (2021), Two-way fixed effects regressions with several treatments. arXiv preprint arXiv:2012.10077.

de Chaisemartin, C. and D'Haultfœuille, X. (2022), Difference-in-differences estimators of intertemporal treatment effects. arXiv preprint arXiv:2007.04267.

de Medeiros, O., Barreto, J., Harris, M., Russo, L. and da Silva, E., 2020. Delivering maternal and childcare at primary healthcare level: The role of PMAQ as a pay for performance strategy in Brazil. *PLOS ONE*, 15(10): p.e0240631.

Facchini L, Tomasi E, Dilélio A. 2018. Quality of primary health care in Brazil: advances, challenges and perspectives. *Saúde em Debate*, 42(sep1): 208–223.

Fardousi, N., Nunes da Silva, E., Kovacs, R., Borghi, J., Barreto, J., Kristensen, S., Sampaio, J., Shimizu, H., Gomes, L., Russo, L., Gurgel, G. and Powell-Jackson, T., 2022. Performance bonuses and the quality of primary health care delivered by family health teams in Brazil: A difference-in-differences analysis. *PLOS Medicine*, 19(7): p.e1004033.

FNS. 2022. *Início - FNS*. [online] Available at: <https://portalfns.saude.gov.br/> [Accessed 1 September 2022].

Golash-Boza, T., 2010. Does whitening happen? Distinguishing between race and color labels in an African-descended community in Peru. *Social Problems* 57(1): 138–156.

Goodman-Bacon, A., 2021. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2): 254–277.

Hone, T., Rasella, D., Barreto, M., Majeed, A. and Millett, C., 2017. Association between expansion of primary healthcare and racial inequalities in mortality amenable to primary care in Brazil: A national longitudinal analysis. *PLOS Medicine*, 14(5): p.e1002306.

Kovacs, R., Maia Barreto, J., Da Silva, E., Borghi, J., Kristensen, S., Costa, D., Bezerra Gomes, L., Gurgel, G., Sampaio, J. and Powell-Jackson, T., 2021. Socioeconomic inequalities in the quality of primary care under Brazil's national pay-for-performance programme: a longitudinal study of family health teams. *The Lancet Global Health*, 9(3): 331-339.

Kovacs, R., Powell-Jackson, T., Kristensen, S., Singh, N. and Borghi, J., 2020. How are pay-for-performance schemes in healthcare designed in low- and middle-income countries? Typology and systematic literature review. *BMC Health Services Research*, 20(1).

Landsberger, H., 1957. *Hawthorne revisited*. Ithaca, N.Y.: Cornell University.

Lima, M. and Prates, I., 2018. Racial Inequalities in Brazil: A Persistent Challenge. *Paths of*

Inequality in Brazil, 113-134

Lovell, P., 2006. Race, gender, and work in São Paulo, Brazil, 1960–2000. *Latin American Research Review*, 41 (3): 63–87.

Macinko, J. and Harris, M., 2015. Brazil's Family Health Strategy — Delivering Community-Based Primary Care in a Universal Health System. *New England Journal of Medicine*, 372(23): 2177-2181.

Macinko, J., Dourado, I., Aquino, R., Bonolo, P., Lima-Costa, M., Medina, M., Mota, E., de Oliveira, V. and Turci, M., 2010. Major Expansion Of Primary Care In Brazil Linked To Decline In Unnecessary Hospitalization. *Health Affairs*. 29(12): 2149-2160.

Matijasevich, A., Victora, C., Barros, A., Santos, I., Marco, P., Albernaz, E. and Barros, F., 2008. Widening Ethnic Disparities in Infant Mortality in Southern Brazil: Comparison of 3 Birth Cohorts. *American Journal of Public Health*, 98(4): 692-698.

Ministry of Health. 2022. *Estratégia Saúde da Família*. [online] Available at: <https://www.gov.br/saude/pt-br/acesso-a-informacao/acoes-e-programas/estrategia-saude-da-familia> [Accessed 20 July 2022].

Paixão, M.J., Rossetto, I., Montovanele, F. and Carvano, L.M., 2010. *Relato´rio anual das desigualdades raciais no Brasil, 2009–2010.* Rio de Janeiro: Editora Garamond.

Pan American Health Organization. 2018. *Health Indicators. Conceptual and operational considerations.* [online] Available at: <https://iris.paho.org/bitstream/handle/10665.2/49056/09789275120057_eng.pdf?sequence=6&isAllowed=y> [Accessed 21 June 2022]

Petersen, L., 2007. Does Pay-for-Performance Improve the Quality of Health Care?. *Annals of Internal Medicine*, 146(7): 538.

Ribeiro, C., 2006. Classe, raça e mobilidade social no Brasil. *Dados*, 49(4):833-873.

Russo, L., Powell-Jackson, T., Maia Barreto, J., Borghi, J., Kovacs, R., Gurgel Junior, G., Gomes, L., Sampaio, J., Shimizu, H., de Sousa, A., Bezerra, A., Stein, A. and Silva, E., 2021. Pay for performance in primary care: the contribution of the Programme for Improving Access and Quality of Primary Care (PMAQ) on avoidable hospitalisations in Brazil, 2009–2018. *BMJ Global Health*, 6(7): p.e005429.

Schlotheuber, A. and Hosseinpoor, A., 2022. Summary Measures of Health Inequality: A Review of Existing Measures and Their Application. *International Journal of Environmental Research and Public Health*, 19(6): 3697.

Schwartzman, L.F., 2007. Does money whiten? Intergenerational changes in reclassification in Brazil. *American Sociological Review* 72(6): 940–963.

Sun, L. and Abraham, S., 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2): 175–199.

Travassos, C. and Williams, D., 2004. The concept and measurement of race and their relationship to public health: a review focused on Brazil and the United States. *Cadernos de Saúde Pública*, 20(3): 660-678.

Unger, J. and De Paepe, P., 2019. Commercial Health Care Financing: The Cause of U.S., Dutch, and Swiss Health Systems Inefficiency? *International Journal of Health Services*, 49(3): 431-456.

Victora, C., Aquino, E., do Carmo Leal, M., Monteiro, C., Barros, F. and Szwarcwald, C., 2011. Maternal and child health in Brazil: progress and challenges. *The Lancet*, 377(9780): 1863-1876.

Wood, C. H. and Carvalho, J. A. M. de., 1994. Categorias do censo e classificação subjetiva de cor no Brasil. *Revista Brasileira De Estudos De População*, 11(1): 3–17.

Wooldridge, J., 2021. Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. *SSRN Electronic Journal*, SSRN 3906345.

Appendix 1

Appendix 1 General Analysis Based on Full Sample 0-64 years

Time	Inequality(1)	Inequality(2)	White(3)	White(4)	Non-White(4)	Non-white(5)
Year 5	-2.38*	-2.49*	1.80*	1.84*	-0.59	-0.65
	(1.34)	(1.35)	(1.07)	(1.05)	(1.33)	(1.35)
Year 4	-2.49	-2.51	1.16	1.14	-1.33	-1.37
	(1.68)	(1.68)	(1.09)	(1.08)	(1.62)	(1.61)
Year 3	-2.58**	-2.58**	0.88*	0.86*	-1.69	-1.72
	(1.14)	(1.13)	(0.47)	(0.47)	(1.16)	(1.14)
Year 2	-1.44	-1.46	0.80	0.77*	-0.65	-0.69
	(0.99)	(0.99)	(0.44)	(0.44)	(0.96)	(0.95)
Year 1	-1.48	-1.46	0.41	0.35	-1.08	-1.11
	(1.14)	(1.15)	(0.41)	(0.41)	(1.10)	(1.12)
Year 0	-0.38	-0.355	0.35	0.33	-0.03	-0.02
	(0.37)	(0.38)	(0.21)	(0.21)	(0.34)	(0.35)
Year -2	0.43	0.42	-0.05	-0.03	0.38	0.39
	(0.36)	(0.38)	(0.24)	(0.24)	0.30	(0.3)
Year -3	0.35	0.29	-0.21	-0.15	0.14	0.14
	(0.71)	(0.73)	(0.48)	(0.49)	(0.56)	(0.57)
Control	No	Yes	No	Yes	No	Yes
Obs.	2699	2699	2699	2699	2699	2699
Switcher	1715	1715	1715	1715	1715	1715

The number of observations and switchers are recorded at Year 0. The column (1) (3) (5) display the estimates without covariates and the column (2) (4) (6) show the estimates with covariates. The standard errors are clustered at municipality level and are displayed in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Appendix 2

Age Subgroups for General Analysis Results

Time	0-4 (Children's Health)			5-19			20-64		
	Inequality	White	Non-white	Inequality	White	Non-White	Inequality	White	Non-White
Year 5	-4.37	5.66	1.30	-1.34	1.29	-0.11	-3.89**	2.12	-1.77
	(7.02)	(5.21)	(6.99)	(0.98)	(0.80)	(0.91)	(1.65)	(1.43)	(1.79)
Year 4	-6.69	3.30	-3.39	-2.14*	1.75**	-0.39	-3.95**	1.48	-2.47
	(11.36)	(4.30)	(11.74)	(1.24)	(0.75)	(0.98)	(1.77)	(1.46)	(1.78)
Year 3	-8.09	-0.02	-8.10	-0.89	0.78	-0.12	-2.91**	0.82	-2.08*
	(6.76)	(2.82)	(6.07)	(1.22)	(0.52)	(1.15)	(1.23)	(0.74)	(1.26)
Year 2	-0.62	-1.18	-1.80	-0.28	0.30	0.025	-0.69	0.59	-0.09
	(3.68)	(2.32)	(2.93)	(1.66)	(0.47)	(1.49)	(0.69)	(0.67)	0.72
Year 1	0.56	-1.86	-1.30	-0.41	0.13	-0.27	-1.11	0.27	-0.85
	(3.30)	(1.97)	(2.77)	(1.63)	(0.44)	(1.43)	0.97	(0.54)	(0.95)
Year 0	-1.28	0.88	-0.40	-0.27	0.18	-0.09	0.04	0.05	0.10
	(1.43)	(0.95)	(1.23)	(0.34)	(0.22)	(0.27)	(0.47)	(0.25)	(0.37)
Year -2	2.36	-0.93	1.43	0.08	-0.27	-0.19	0.82*	-0.22	0.60*
	(1.91)	(1.30)	(1.97)	(0.48)	(0.32)	(0.28)	(0.44)	(0.29)	(0.33)
Year -3	5.68	-4.31	1.38	0.59	-0.69	-0.10	-0.07	-0.25	-0.32
	(4.44)	(2.78)	(3.17)	(0.52)	(0.45)	(0.46)	(0.66)	(0.47)	(0.53)
Obs.	2656	2656	2656	2609	2609	2609	2643	2643	2643
Switcher	1704	1704	1704	1684	1684	1684	1703	1703	1703

The number of observations and switchers are recorded at Year 0. The standard errors are clustered at municipality level and are displayed in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Appendix 3

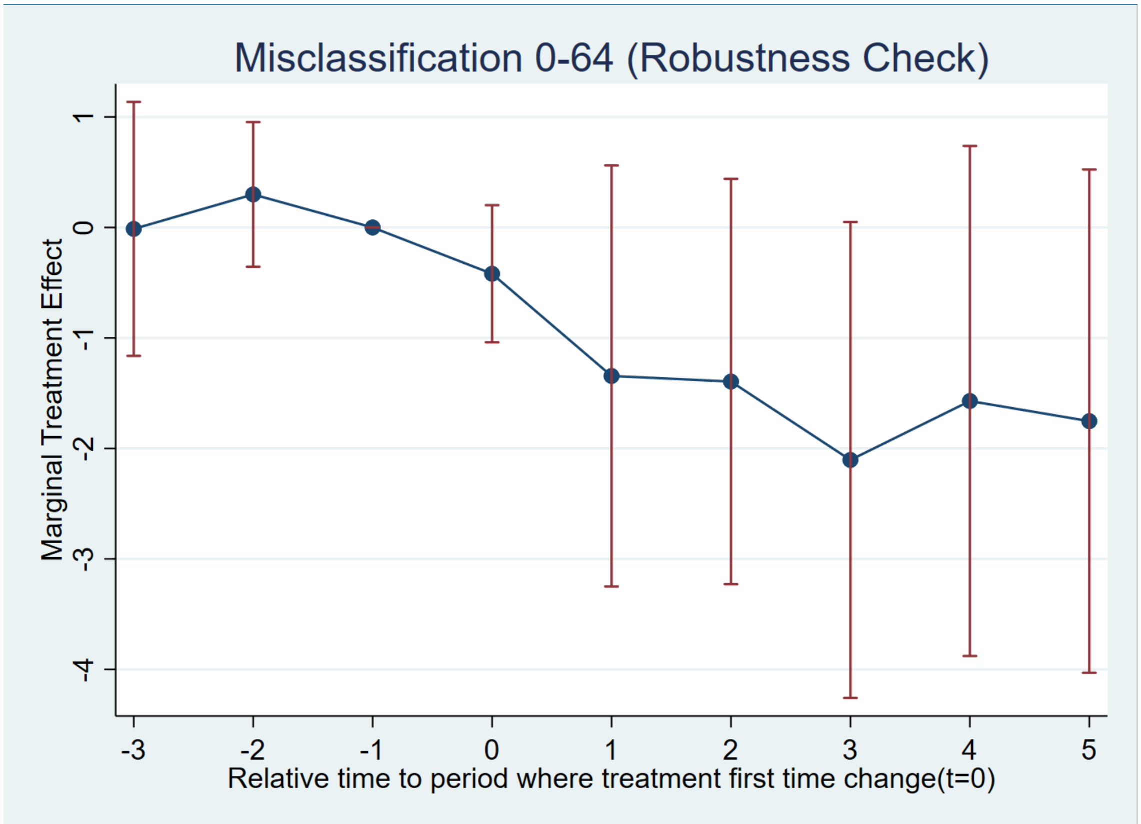
Specific Analysis Result

Time	Woman's Health			DHTL		
	Inequality	White	Non-White	Inequality	White	Non-White
Year 5	-0.022	0.017	0.039	-0.62	0.16	-0.45
	(0.033)	(0.025)	(0.030)	(0.40)	(0.24)	(0.42)
Year 4	-0.021	0.028	0.048	-0.55	0.15	-0.40
	(0.035)	(0.027)	(0.036)	(0.38)	(0.19)	(0.36)
Year 3	-0.020	0.001	0.020	-0.27	0.18	-0.10
	(0.020)	(0.017)	(0.021)	(0.19)	(0.12)	(0.19)
Year 2	-0.023	0.029*	0.026	-0.18	0.01	-0.17
	(0.020)	(0.016)	(0.016)	(0.19)	(0.11)	(0.16)
Year 1	-0.021	-0.001	0.020	-0.28*	-0.12	-0.29*
	(0.018)	(0.013)	(0.015)	(0.17)	(0.09)	(0.17)
Year 0	-0.016	-0.007	0.009	-0.10	0.05	-0.052
	(0.013)	(0.009)	(0.009)	(0.08)	(0.04)	(0.071)
Year -2	0.019	0.016	-0.002	0.04	0.065	0.10
	(0.012)	(0.010)	(0.010)	(0.07)	(0.05)	(0.07)
Year -3	-0.005	-0.018	-0.013	0.02	-0.04	-0.02
	(0.020)	(0.017)	(0.016)	(0.13)	(0.10)	(-0.14)
Obs.	6784	6784	6784	2668	2668	2668
Switcher	4787	4787	4787	1696	1696	1696

The number of observations and switchers are recorded at Year 0. The standard errors are clustered at municipality level and are displayed in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Appendix 4

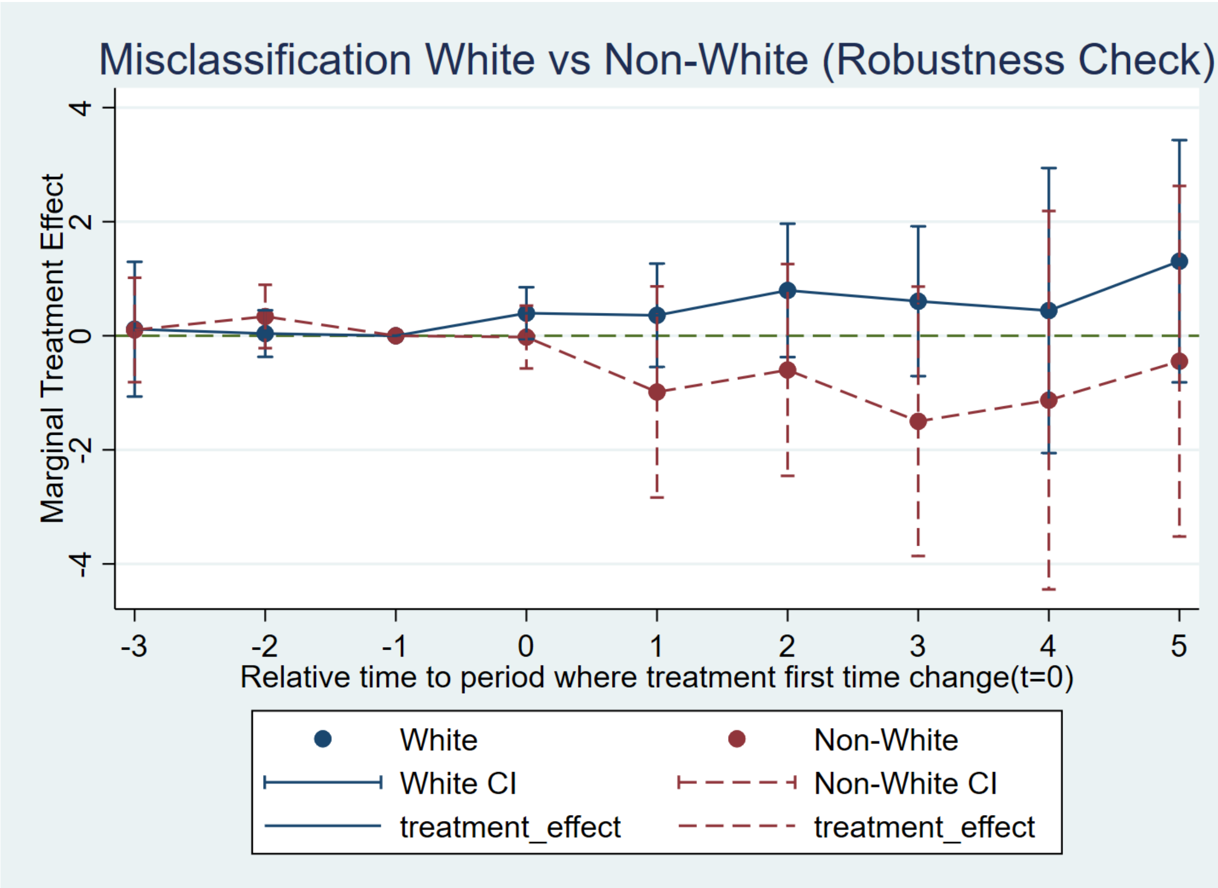
Race Misclassification Robustness Check Based on Full Sample (0-64)



Appendix 4 displays the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 0-64 years for inequality constructed with ACSC hospitalization rate after re-assigning non-white patients to white. The standard errors are clustered at municipality level.

Appendix 5

Racial Subgroup Trend Based on Full Sample (0-64): Race Misclassification Robustness Check



Appendix 5 presents the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 0-64 for ACSC hospitalization rate of white (Blue) and non-white (Red) after re-assigning non-white patients to white. The standard errors are clustered at municipality level.

Appendix 6

Robustness Check Results

Time	Misclassification			Racial Discrimination
	Inequality	White	Non-White	Inequality
Year 6	-	-	-	-0.07
	-	-	-	(0.12)
Year 5	-1.75	1.30	-0.45	-0.08
	(1.16)	(1.08)	(1.57)	(0.12)
Year 4	-1.57	0.44	-1.13	0.03
	(1.17)	(1.27)	(1.69)	(0.11)
Year 3	-2.10*	0.61	-1.50	-0.01
	(1.10)	(0.67)	(1.20)	(0.59)
Year 2	-1.39	0.79	-0.60	0.01
	(0.94)	(0.60)	(0.95)	(0.09)
Year 1	-1.35	0.36	-0.99	0.02
	(0.97)	(0.46)	(0.94)	(0.06)
Year 0	-0.42	0.40*	-0.02	0.00
	(0.32)	(0.23)	(0.28)	(0.04)
Year -2	0.30	0.04	0.34	0.03
	(0.33)	(0.21)	(0.28)	(0.04)
Year -3	-0.01	0.11	0.10	-0.03
	(0.59)	(0.60)	(0.47)	(0.06)
Obs.	2699	2699	2699	4140
Switcher	1715	1715	1715	2838

The number of observations and switchers are recorded at Year 0. The standard errors are clustered at municipality level and are displayed in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

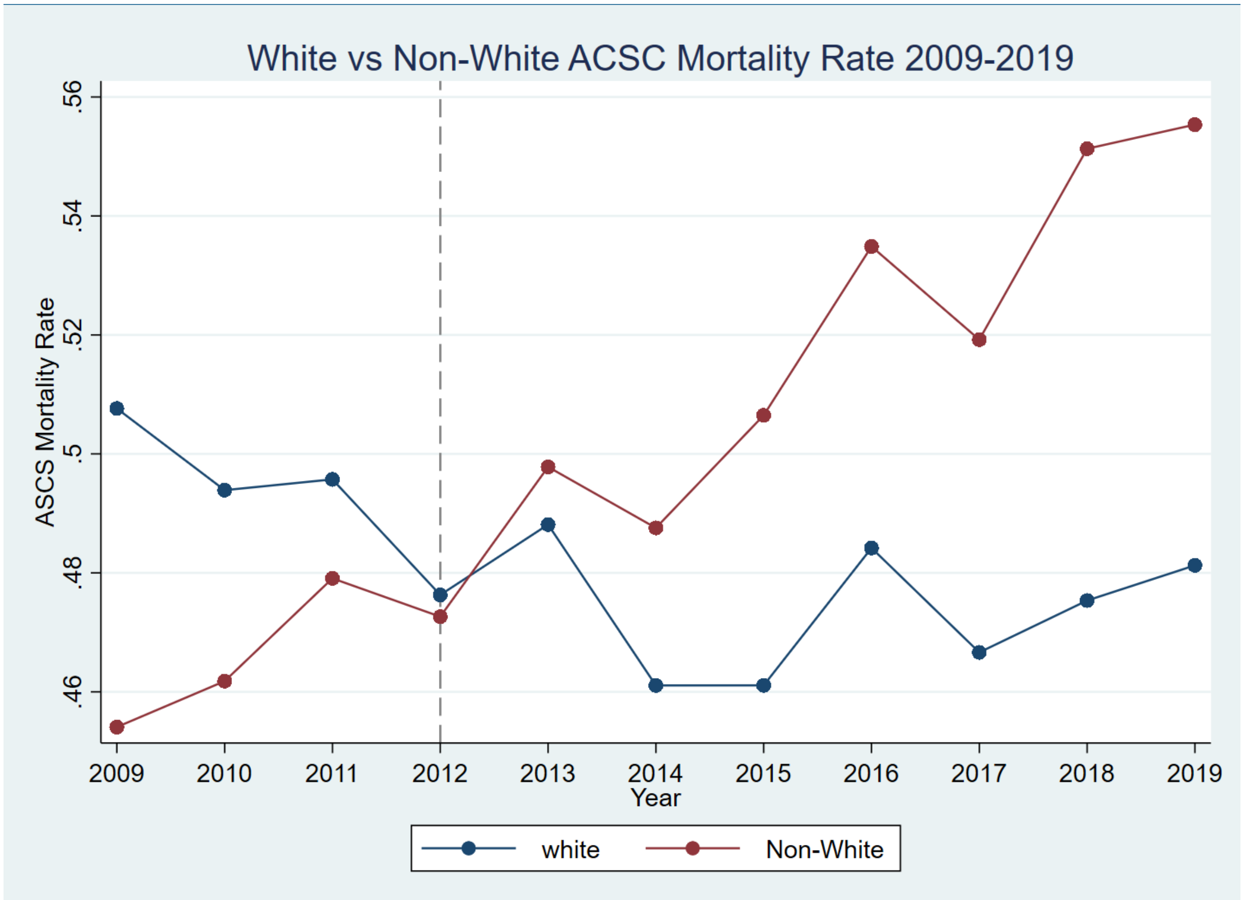
Appendix 7

Racial Discrimination Robustness Check Based on Full Sample (0-64)

Appendix 7 displays the trend of the de Chaisemartin and D'Haultfoeuille (2022) estimates with 95% confidence intervals based on the 0-64 years for inequality constructed with ACSC mortality rate. The standard errors are clustered at municipality level.

Appendix 8

ACSC Mortality Rate Trend for 2 Racial Groups from 2009 to 2019



Appendix 8 presents the trend of ACSC mortality rate of white (Blue) and non-white (Red) from 2009 to 2019 based on the 0-64 years sample. The first round of the PMAQ started in November 2011.

Appendix 9

Placebo Test for No Anticipation Hypothesis

	Placebo Year -2	Placebo Year -3		Placebo Year -2	Placebo Year -3
0-64			**20-64**		
Inequality	-0.42	0.27	Inequality	-0.82*	0.78
	(0.38)	(0.39)		(0.44)	(0.68)
White	0.03	-0.33	White	0.22	-0.37
	(0.24)	(0.27)		(0.29)	(0.27)
Non-white	-0.39	-0.07	Non-white	-0.60	0.42
	(0.30)	(0.38)		(0.33)	(0.32)
0-4 (Children's Health)			**Women's Health**		
Inequality	-2.36	-0.047	Inequality	-0.019	0.006
	(1.91)	(2.17)		(0.012)	(0.011)
White	0.93	-0.10	White	-0.016	0.018
	(1.29)	(1.28)		(0.010)	(0.009)
Non-white	-1.43	-0.15	Non-white	0.002	0.012
	(1.97)	(1.88)		(0.010)	(0.008)
5-19			**DHTL**		
Inequality	-0.08	0.23	Inequality	-0.036	0.000
	(0.48)	(0.31)		(0.071)	(0.085)
White	0.27	-0.51	White	-0.065	0.028
	(0.32)	(0.26)		(0.050)	(0.06)
Non-white	0.19	-0.28	Non-white	-0.100	0.028
	(0.28)	(0.27)		(0.067)	(0.068)

The standard errors are clustered at municipality level and are displayed in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Appendix 10

Placebo Test for Heterogeneous Treatment Effect

	Num of Positive Weights	Num of Negative Weights
0-64		
Inequality	9915	7563
White	9915	7563
Non-white	9915	7563
0-4 (Children Health)		
Inequality	10694	8017
White	10694	8017
Non-white	10694	8017
5-19		
Inequality	10749	7894
White	10749	7894
Non-white	10749	7894
20-64		
Inequality	10005	7738
White	10005	7738
Non-white	10005	7738
Women Health		
Inequality	21696	15408
White	21696	15408
Non-white	21696	15408
DHTL		
Inequality	10623	7868
White	10623	7868
Non-white	10623	7868

Appendix 10 displays the placebo test for the sign of the assigned weight for a specific municipality for a specific year when calculating the average treatment effect with the traditional difference-in-differences method. The existence of the negative weights implies the poor performance of the traditional DiD.