ISER Working Paper Series

Productivity Effects of Dengue in Brazil

Sonia Bhalotra
University of Essex

Gabriel Facchini
Universitat Autonoma de Barcelona

Aline Menezes
Fundacao Getulio Vargas

Rudi Rocha
Fundacao Getulio Vargas

No. 2019-04
May 2019

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

University of Essex

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

www.iser.essex.ac.uk

Non-Technical Summary

This paper investigates the economic costs of dengue outbreaks in Brazil, and the mitigating impacts of statutory sick pay and public health clinics. Rather than using the raw variation in dengue, we filter the data to isolate variation in dengue generated by random variation in temperature and humidity, using a particular functional form that emerges from an epidemiological model. To allow that climatic conditions may have direct impacts on productivity, we difference between pre and post dengue periods, leveraging the fact that dengue has only re-appeared in Brazil since 2007.

We find that dengue outbreaks are associated with a loss of hours of work and earnings in the non-formal sector, especially among women. We argue that the gender difference is likely to arise on account of women taking time off work not only to care for themselves but also for other members of the household who are sick. Cash transfers like Bolsa Familia are designed to address structural poverty rather than temporary shocks to earnings capacity. Among formal sector workers, we find no loss in work hours or earnings but a rise in state subsidized sick-pay (a form of welfare). These results are reinforced using longitudinal data that allow us to identify individual (rather than regional) incidence of dengue.

Using a more localised but geo-coded data set, we find that the opening of primary health care facilities leads to improvements in income among people on the welfare payments register and, again, most clearly among women working in informal sector jobs (that do not insure against sickness).

Our findings contribute to a literature concerned with identifying the causal impact of health shocks, health provision and income protection on income and poverty. They also contribute to a different branch of work concerned with impacts of climate change on infectious disease prevalence insofar as global warming is predicted to lead to more frequent and more widespread dengue outbreaks.

Our research benefited from the use of several administrative data sets linked at the individual or regional level, alongside survey data. The quality of the data are high. We obtained dengue infection counts based on clinical evaluation rather than self-diagnosis, a census of hospitalizations by cause, daily data on humidity and temperature, and welfare registers containing longitudinal data on the labour market status and earnings of the universe of welfare recipients, for a fraction of whom we have complete linkage to individual dengue records identifying the exact date at which an individual contracted dengue. We also obtained administrative data on sick pay and on cash transfers under Bolsa Familia. For the state of Rio de Janeiro, we have administrative data that record the exact date of opening of primary care facilities (GP clinics) including the exact location (post code), allowing us to study how outcomes in a well-defined catchment area evolve after clinic opening. These data are matched to individual employment and earnings data in the administrative welfare registers which also contain postcode.

Productivity Effects of Dengue in Brazil

Sonia Bhalotra, Gabriel Facchini, Aline Menezes, Rudi Rocha*

May, 2019

Abstract

Although understanding the role of health in driving labor market outcomes is a matter of great importance, it has proven difficult to isolate this effect due to empirical challenges and a lack of compelling sources of identification. We obtain causal estimates of the effect of health on income and welfare dependency through two different channels: a negative health shock (dengue outbreak) and a positive health shock (opening of a health-care facility). To do this, we rely on instrumental variables and difference-in-difference methods, as well as on novel datasets. We find that dengue outbreaks lower the average working hours and income. This effect is particularly high for low-income individuals, but conditional cash transfer programs can insulate them from this shock. On the other hand, the opening of a new health-care facility in a families catchment area rises family per capita income and employment. All together, this evidence suggest that health shocks are an important part of income, poverty and welfare dependency.

*Bhalotra: University of Essex (email: srbhal@essex.ac.uk); Facchini: Universitat Autònoma de Barcelona (email: gabriel.facchini@uab.cat); Menezes: Fundação Getulio Vargas (email: alinemsmenezes@gmail.com); Rocha: Fundação Getulio Vargas (rudi.rocha@fgv.br)

1 Introduction

There is a large literature documenting a positive correlation between health and income, but the direction of causation that explains most of this correlation remains largely unknown (Deaton, 2003; Weil, 2014). Indeed health tends to improve with income, as food intake increases in quantity and quality, more resources can be devoted to health care and stress about basic needs fulfillment is reduced. However health can also affect income in return, notably through workers' productivity, as it may affect their absenteeism, cognitive capacities but also incentives to invest in physical, human or institutional capital. This bi-directional relationship between health and income, together with a lack of compelling sources of identification, has made it difficult to isolate the role of health in driving labor market outcomes (Currie and Madrian, 1999).

In this paper we revisit this issue and try to quantify the chain of causation running from health to income and, ultimately, welfare dependency. To this end, we explore two avenues: first, we explore the effects of negative health shocks such as infectious diseases' outbreaks; second, we look at positive health shocks provided by the implementation of health-care programs.

In the first part, we use dengue outbreaks as our health measure. Among all mosquito-borne viral diseases in the world, dengue is the one experiencing the fastest spreading: incidence has increased 30-fold in the last 50 years (World Health Organization, 2009). Although previously confined to tropical regions, in recent years dengue has spread into temperate zones.[1] A large part of this expansion is due to the rise in air travel and trade (Bloom et al., 2006). There are several papers in the epidemiological literature that aim at quantifying the economic burden of dengue outbreaks (see Beatty et al. (2011) for a systematic review of the literature, and Undurraga et al. (2015) and Shepard et al. (2016) for more recent examples).[2] However, this literature ignores important issues such as measurement error and endogeneity, and are based on small sample studies.[3] In contrast, this article improves on previous studies by estimating a model using data for the whole population of Brazil and explicitly dealing with the endogeneity of regressors and measurement error in reported dengue cases.

We first do an analysis of the relationship between dengue cases and labor market outcomes at the metropolitan-region level using information from Brazil's labor force survey (Pesquisa Mensual de Emprego) and exploiting weather variations as an exogenous local factor that greatly affects the population dynamics of mosquitoes and hence, that of mosquito-borne viruses like dengue.

[1] Messina et al. (2014) document the global distribution of confirmed cases of dengue virus from 1943 to 2013.

[2] One of the most comprehensive and cited study in this literature is Suaya et al. (2009). Using a sample of ambulatory and hospital dengue cases for 8 countries in the world, these authors estimate that an average dengue episode leads to a 9.9 loss in working days. They calculate the productivity cost of dengue multiplying this by the national daily minimum wage and national estimate of dengue cases.

[3] When they do tackle the issue of measurement error, they use an expansion factor to adjust for under-reporting in national notified cases. These expansion factors cover a wide range: between 1.6 and 3.2 for hospital cases and between 10 and 27 for ambulatory cases (Suaya et al., 2009).

We use as instrument for local dengue count a "Dengue Suitability Index", developed by Obolski et al. (2018), which depends on local temperature and relative-humidity conditions. The exclusion restriction is satisfied as long as these weather variables do not have a direct effect on our labor market outcomes, or they do so through a different functional form. We show in the paper that our instrumental variable has no direct effect on outcomes during years with no dengue outbreaks, but a statistically and economically significant effect during years with outbreaks. The drawback of this analysis is that we can only obtain an intention-to-treat effect since we cannot identify individuals that have dengue.

However, for the period of January to April of 2017, we have identified dengue cases that we then match to Brazil's registry of welfare recipients (CadUnico) -which has information on income and other labor market outcomes. From CadUnico, we first select families who updated their records both in 2016 and 2017, and then draw all individuals that belong to those families. Records between the dengue reported cases registry and CadUnico are then matched using individuals names, exact date of birth, gender and municipality of residence. We then exploit different samples, the timing of the disease, and rely on a difference-in-differences (DiD) specification in order to balance treatment *vs* control groups, so to recover the causal effect of dengue on income and welfare. Our results suggest that, when the head of the household is hit by dengue, the family income per capita drops by approximately 28%. This result is robust to the inclusion of different fixed effects and sample selection. Income doesn't seem to be affected when the spouse or children are hit by dengue. Finally, individuals who are beneficiaries of Brazil's main conditional cash transfer program (Programa Bolsa Familia) are insulated from the detrimental effects of dengue on family income. However, we do not serve families taking up the program once they are hit by the dengue.

Finally, the second part of the paper uses the expansion of health care facilities as a positive health shock. There exists a wealth of evidence that programs aiming at expanding healthcare services to populations that initially did not have access were successful at improving health outcomes, as measured by mortality rates or birth outcomes (Rocha and Soares, 2010; Bhalotra et al., 2019; Bailey and Goodman-Bacon, 2015; Goodman-Bacon, 2018). We add another channel through which society benefits from this facilities, the potential effect of improved health on income and welfare dependency.

We use as treatment opening of new Family Health Programme (PSF) clinics, a project from the Brazilian Ministry of Health that provides preventive and basic care through interventions at the community level. These clinics are assigned a catchment-area to which they have to provide service. We geocoded each new opening and matched it to geocoded families in CadUnico within their catchment area. Using the facilities' sequential implementation across neighborhoods in Rio de Janeiro during the first half of 2010's together with a difference-in-difference strategy, we find that the opening of a health facility increases employment and incomes for individuals in the treated catchment area. These findings are robust to the inclusion of different type of time and space controls specifications.

The remainder of the paper is structured as follows. In Section I we summarize the more recent economic literature on the effect of health and income. In Section II we analyze the effect of dengue outbreaks on income and welfare dependency outcomes. In Section III we use health-care facility opening as treatment and look at its effect on income and welfare dependency. We conclude in Section IV.

2 Related Economic Literature

As previously mentioned, although understanding the role of health in driving labor market outcomes is a matter of great importance, it has proven difficult to isolate this effect due to empirical challenges and a lack of compelling sources of identification (Currie and Madrian, 1999).[4] Not surprisingly, most recent studies on this issue have use data from developed countries (US and European countries), where it is possible to access administrative data and connect health shock measures (e.g. hospitalizations) with outcomes (e.g. income) at the individual level. This data allows them to reduce endogeneity concerns by performing event studies type of analysis -where researchers can check for pre-trends in outcomes before the shock took place. Based on their revision of this literature, Prinz et al. (2018) conclude that health shocks are an important driver of labor market transitions, but highlight that the magnitude of the effects depend on the institutional environment in each country.

We contribute to this literature in two ways. First, we use as health shocks cases from an infectious disease that con become epidemic. Previous studies look at health shocks that are relatively important for the individual worker but not for the society (e.g. cancer, HIV, some physical disability, etc). Instead, dengue shocks are, in most cases, not severe from a patient's view but have the potential to affect a large fraction of the population in a short period of time. One exception is Marinescu (2014), who uses regional variation in HIV prevalence for several countries in Africa between two years (1990 and 2000) to look at its effect on firms and workers behavior. In order to reduce measurement error and omitted variable bias concerns, she uses regional circumcision rates as instrument for HIV prevalence. She finds no significant impact of HIV on labor productivity and a small effect on hours worked (going from 9% to 18% HIV prevalence would decrease weekly hours worked by 3 hours). However, a higher skill premium is associated with higher HIV prevalence -which she relates to higher mortality and reduced labor supply.

Second, as mentioned before, studies looking at the effect of health shocks using data from devel-

[4] In a recent study Adda (2016) analyze the other direction of this relationship and find that viral diseases spread faster during economic expansions in France.

oping countries are rather scarce. However, the effects are probably of bigger magnitude in settings with a large informal sector: these workers are not enrolled in any sickness absence scheme and will not get paid for taking days off due to sickness. Gertler and Gruber (2002) study the effect of major illnesses on income and consumption using panel data from Indonesia. Taking advantage of questions individuals' self-ratings of ability to engage in specific activities, the authors construct a measure of physical ability to perform activities of daily living (ADLs). Using a first-difference estimator, they find that if the household-head move from completely healthy to completely sick, her hours of work per week would fall by almost 31 hours. Wagstaff (2007) look at the effect of three different health shocks (death of a relative, hospitalization of a household-member and drop in the head of the household's body mass index) on income and consumption at the household level. Using cross-sectional data from the Vietnam Living Standards Survey together with health shocks from a preceding period as treatment, they find evidence of a negative impact of some health shocks on income and consumption. However, these papers cannot totally rule out the existence of confounding factors at the individual level that change with time and are correlated with both health and income. We expand on this work by looking at an understudied infectious disease whose characteristics -mosquito-borne disease with unexpected shocks- allow us to obtain causal estimates of its effect on the labor market.

A closely related literature looks at the association between health insurance (as opposed to health) and labor market outcomes. A large part of these studies use policies giving health insurance to previously uninsured workers and analyse their behavior and outcomes at the time of the change. Instead, few studies have focused in understanding the beneficial effect that health insurance can have in breaking the connection between health shocks and labor market outcomes. One exception is Baicker et al. (2014), who compare individuals the year after the Oregon 2008 health-insurance-lottery in terms of employment and earnings. They find no statistically significant difference between winners and losers of the lottery. Using the same experiment, Finkelstein et al. (2012) find that a year later the treatment group had higher health care utilization and lower out-of-pocket medical expenditures. Mazumder and Miller (2016) arrives to a similar conclusion using major health care reform in Massachusetts. These has also been found to be the case in developing countries (Limwattananon et al., 2015; Miller et al., 2013).

Finally, this paper also speaks to a rapidly growing literature in economics looking at climate change and its effect on human activities.[5] Previous studies have found compelling evidence that temperature can have a direct effect on productivity (Graff Zivin and Neidell, 2014; Heal and Park, 2015; Zhang et al., 2018), as well as on health (Deschênes and Greenstone, 2011; Kudamatsu et al., 2012). The results in this article contribute to this literature by adding the previously unexplored role of weather on disease spreading as a complementary channel that links weather to labor market outcomes.

[5]See Dell et al. (2014) for a comprehensive review of the literature

3 A Negative Health Shock: Dengue

3.1 Epidemiological Background

Dengue fever (DF) is caused by any of four closely related viruses, or serotypes: dengue 1-4. The viruses are transmitted by female mosquitoes, mainly of the species Aedes aegypti and, to a lesser extent, Aedes albopictus. These mosquitoes can be found throughout the tropics, with local variation in risk influenced by weather and urbanization -dengue is primarily and urban disease.

Symptoms of infection usually begin 4–7 days after the mosquito bite and typically last 3–10 days. They include high fever, severe headache, and joint and muscle pain. These symptoms generally dissipate without medical treatment and present no lasting effects. In some people, however, dengue can cause small blood vessels to leak, causing Dengue Hemorrhagic Fever (DHF), which can lead to death. It is treated with therapies that ameliorate the symptoms until the patient's immune system can overcome the disease.

There are still no vaccines to prevent infection with dengue virus, but early recognition and care treatment can significantly lower the risk of medical complication and death. Individuals are encouraged to eliminate potential mosquito habitats- any container that can accumulate clean water. Infection with one serotype does not protect against the others. Furthermore, people with sequential infection are at greater risk for developing DHF and dengue shock syndrome.

Because dengue infections are climate sensitive, recent studies focus on understanding how changes in climatic factors may affect the potential spread of the disease. Using population and climate change projections for 2085, Bhatt et al. (2013) estimate that about 50-60% of the (projected) global population would at risk of dengue transmission -compared to 35% in the absence of climate change. In a similar study, Liu-Helmersson et al. (2014) estimate a measure of local dengue-epidemic-potential (DEP) based on the temperature and diurnal temperate range dependence of the mosquito. They then use these parameters and temperature predictions until 2099 and find and increasing trend of global DEP for temperate regions, concentrated in the Northern Hemisphere.

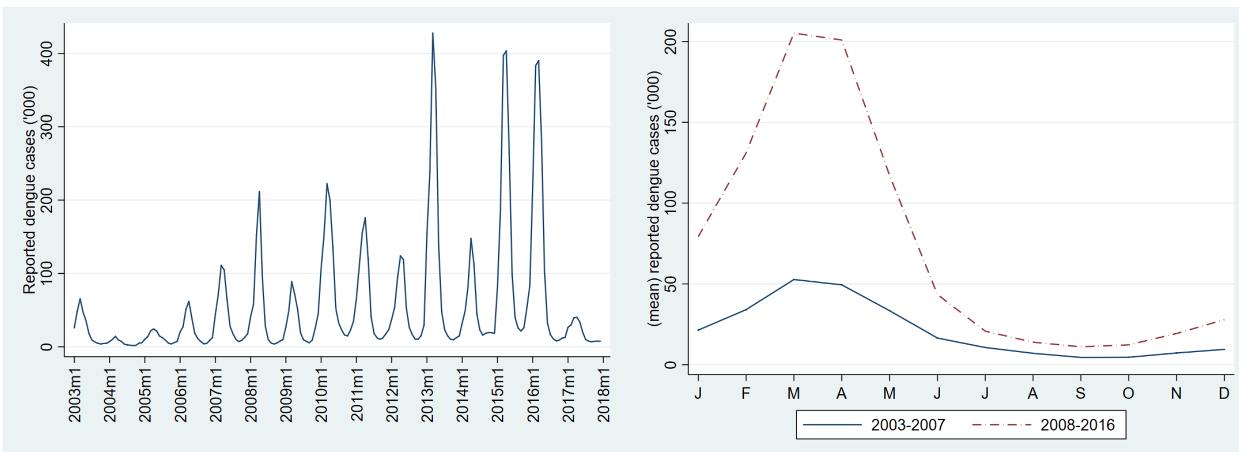
3.1.1 Dengue in Brazil

In 2016 there were slightly more than 3.2 million dengue cases reported worlwide, and about half of them (1.65 million) took place in Brazil.[6] This makes Brazil a particularly interesting setting to study the effects of dengue. Figure 1a shows the time series for dengue reported cases. The yearly spike in the first semester is evidence of dengue's seasonlity. This is more evident in Figure 1b,

[6]Source: (WHO) Dengue and Severe dengue: Key facts (`http://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue`); and (PAHO) Health and Information Platform for the Americas (`http://www.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-nacional-en/252-dengue-pais-ano-en.html`).

which takes the average for each month across the sample years. The first semester of the year, the period of highest temperature and rainfall levels in the largest part of Brazil, is particularly sensible to dengue outbreaks. Figure 2 shows the dengue incidence per 10,000 people across municipalities in Brazil in 2013 -dengue outbreak year. Figure 2a shows incidence for the month of March (high incidence), and Figure 2a for September (low incidence). It is interesting to see how municipalities with high incidence are concentrated in the middle part of the country (Tropical Central Brazil). In Table 1 we show the distribution of dengue incidence for two samples: all months of the year, and only for March thorough May, those of high dengue-season.

Figure 1: Reported dengue cases - Brazil 2008-2016



(a) (b)

Table 1: Distribution of dengue incidence (cases per 10,000)

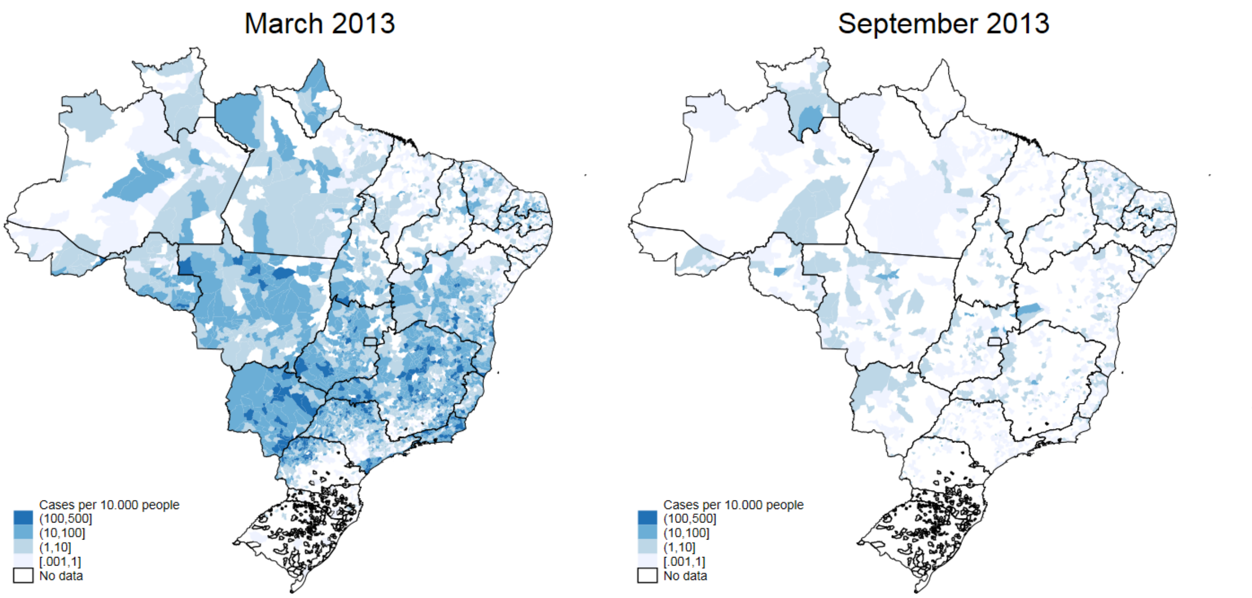
	mean	sd	min	max	p25	p50	p75	p90	p99
All months	2	1.22	0.00	1,423	0.00	0.00	0.2	2.8	41.6
March-May	4.5	1.86	0.00	1,423	0.00	0.00	0.17	9.2	78.6

3.1.2 Dengue vs. Malaria

Malaria, together with dengue, are the two most common arthropod-borne diseases. Differently from dengue, approximately 90% of both Malaria cases and deaths occurred in Africa.

Alhtough the ratio of death to cases is greater for Malaria, in both cases patients present similar symptoms. In communities where dengue fever and malaria occur at some frequency in adults, it

Figure 2: Dengue incidence across municipalities and seasons



(a) (b)

may be difficult to discriminate patients presenting dengue fever symptoms from patients presenting malaria attacks.

In our paper we use dengue reported cases, hence there is a potential risk of counting malaria cases as dengue. However, while dengue is primarily found in urban centers, Malaria cases in Brazil during our period of analysis were very few and restricted to the Northwest part of the country, in the Amazonas (Panamerican Health Organization, 2014). Furthermore, our study of the labor market uses survey data for 6 metropolitan-regions -regions form now on-, all of which are located in the east side of the country (see section 3.2.2).

3.2 Macro Effects

3.2.1 Data Construction and Empirical Framework

3.2.1.1 Estimating Equation

Motivated by our discussion above, the empirical analysis focuses on how dengue infections affect labor market outcomes. The main estimating equation is presented below:

$$\mathrm{y}_{it} = \beta\, Dengue_{it} + \lambda_i + year_t + month_t + \epsilon_{it} \tag{1}$$

where y_{it} is our labor market outcome of interest for metropolitan-region i and period t. $Dengue_{it}$ is the inverse hyperbolic sine (IHS) transformation of the number of reported dengue cases.[7] To control for secular variation in outcomes, we also include yearly and monthly dummies; and λ_i represent metropolitan-region fixed effects.

Despite using a demanding fixed effect estimator in Equation (1), two threats to identification remain. Our primary concern is omitted variable bias, the omission in our model of time-varying variables that are correlated with ϵ and *Dengue*. For example, Adda (2016) find that economic expansions increase the spread of three major viral diseases: influenza, gastroenteritis and chickenpox. If this is true for dengue as well, then OLS estimates of β would be biased downwards. A secondary concern is is measurement error in dengue cases, which the literature has already found to be substantial.[8] We discuss in Appendix B.4 possible forms of the measurement error and how they affect our estimates. We address both of these issues by using instrumental variables.

One important factor of the population dynamic of mosquito-borne viruses is that they depend on the population dynamics of their vector species. There are many time-invariant factors that dictate mosquitoes' absolute population sizes (carrying capacity): altitude, climate, population density, water reservoirs and others. However, natural climate variations drive the seasonal oscillations (Brady et al., 2013), and hence we can use changes in weather as a source of exogenous variation in the mosquito population.

We draw on Obolski et al. (2018), who wrote an R code (MVSE) to compute a climate-driven, mathematical model of mosquito-borne viral transmission based on ordinary differential equations. Given local daily conditions of average temperature and relative humidity, and some virus and vector dynamic parameters, this package estimates a mosquito-borne viral suitability index -which for the case of dengue we will call a Dengue Suitability Index (DSI).[9] We then use this index as an instrumental variable for dengue cases, as shown in Equation (2):

$$Dengue_{it} = \pi \, \mathrm{DSI}_{it} + \lambda_i + year_t + month_t + u_{it} \tag{2}$$

Figure 3 shows aggregate dengue cases and DSI for our PME sample period. We can see that DSI has a strong seasonal component, with a spike at the beginning of each year and a low point in the months of July/August. Dengue cases also show this seasonal components, but with some years with very low dengue and others very high. This relationship between DSI and dengue cases survives the inclusion of region and time controls: current DSI is significantly correlated to dengue

[7] Formally, the inverse hyperbolic sine transform of some variable, y, is $ln(y + \sqrt{(y^2 + 1)})$.

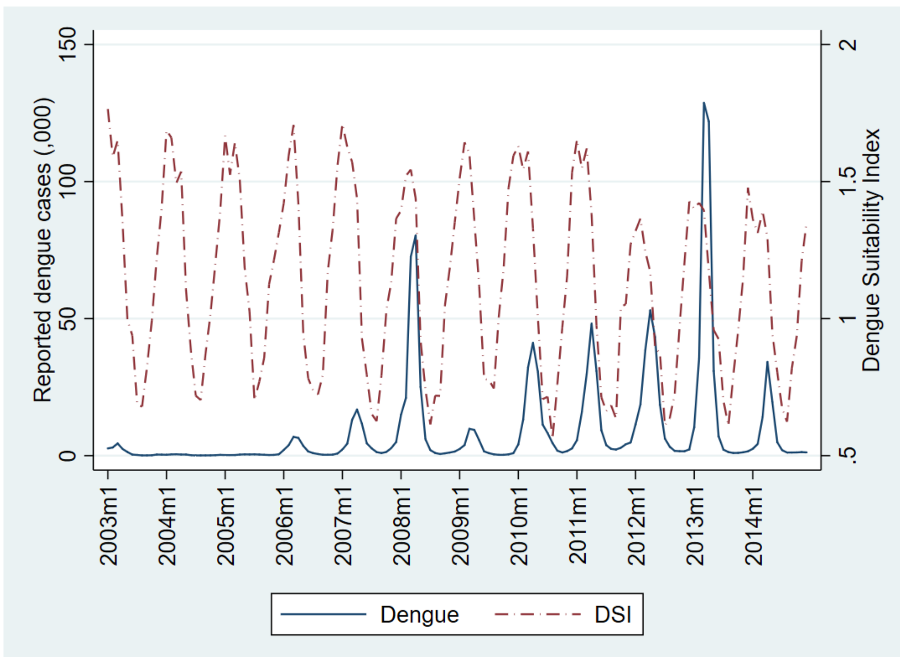
[8] For example, Silva et al. (2016) estimate that, in Brazil, only 1 in every 12 cases are reported during dengue season. This ratio is 1/17 for low dengue transmission periods.

[9] Human and mosquito parameters were set as specified in the "Full Methods and Data Description" appendix in Obolski et al. (2018) for all municipalities, and used Gaussian distributed priors and simulated samples of size 1000.

cases at over 95 percent confidence across specifications (Table 2). Note that once we include month FE (column 2 vs. column 1) the point estimates drops by almost 50% and the F-statistic is much lower too.[10]

Figure 3 also shows that, for these six metropolitan-regions, there were very few dengue cases between 2003 and 2007. Only from 2008 onward there are important dengue outbreaks every year (but 2009). For this reason, our the main part of our analysis will use data from 2008 until 2014, and include the period pre-dengue (before 2008) as placebo in following robustness checks.

Figure 3: Time series of dengue and DSI



3.2.2 Data

Dengue Data: Weekly data on the number of notified dengue cases (mostly by clinical evaluation, not laboratory confirmed) for the period 2000-2017 by municipality are available from the Notifiable Diseases Information System (Sistema de Informação de Agravos de Notificação - SINAN). The general practitioner (GP) who suspects a patient is infected with dengue is in charge of notifying it. This administrative database serves as a reference for epidemiological surveillance activities. As mentioned before, dengue cases are highly under-reported. We describe this data in greater length

[10]We do not include weather controls in our main regressions because there is a very high correlation between DSI and temperature and humidity. Hence by including these controls we loose the power of our instrument. See Appendix B.3 for a detail explanation of how each weather variable enters in the DSI estimation, and their relationship. To lower concerns of capturing direct effects of weather on dengue we will perform a serious of robustness checks in the next section.

Table 2: DSI and Dengue (First-Stage)

	Dengue			
	(1)	(2)	(3)	(4)
DSI	0.1840***	0.1005***	0.1353***	0.1005***
	(0.0195)	(0.0265)	(0.0293)	(0.0267)
Observations	648	648	648	648
IV F-stat	88	14	21	14
Year FE	Yes	Yes	No	Yes
Month FE	No	Yes	No	Yes
Year*Month FE	No	No	Yes	No
Region time-trend	No	No	No	Yes

Notes: Each coefficient represents a separate regression of the number of GP dengue reported cases on the estimated Dengue Suitability Index (DSI). The sample includes all 6 metropolitan-regions represented in PME. Both Dengue (dependent variable) and DSi (independent variable) are standardized as Z-scores so coefficients are interpreted as standard deviation change in dengue cases associated with a 1 sd increase in DSI. All models include metropolitan-region fixed effects. Robust standard errors in parentheses, significance ***p<0.01, **p<0.05, *p<0.1

in Appendix A.2.

Weather Data: We use daily weather data (average daily humidity and temperature) from over 3,600 weather stations from Brazil's National Institute of Meteorology (INMET). We obtain municipality-specific weather measures following an interpolation method that uses as weights the inverse of the distance between municipalities' geographic centroid and all weather stations.[11]

An alternative approach would be to use an existing weather database at monthly frequency. The most commonly used datasets of global temperature and precipitation data are: (i) Willmott and Matsuura, at the University of Delaware, and (ii) Climatic Research Unit (CRU) at the University of East Anglia. However, daily weather station data allows us to calculate a daily DSI that can then be aggregated at higher levels, therefore retaining some within month variability in the index. In any case, the correlation between our and CRU's time series for temperature is very high (.96).

Labor Market Data: We use data from the PME (Brazil's Labor-Force Survey) from 2008 until M9-2014. It is representative at the level of 6 metropolitan regions (Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo and Porto Alegre).[12]

[11] Xavier et al. (2016) find that inverse distance weighting and angular distance weighting are the methods that produce the best results for the Brazilian case out of 6 different interpolation methods.

[12] Figure B.1 shows a time series of dengue reported cases for every 1,000 inhabitants by metropolitan-region in our PME sample. We can observe that the two regions in the north (Recife and Salvador) show more years of dengue

We restrict the sample to people between 15 and 65 years old, and dropped public employees, employers, and people with no remuneration (e.g. temporary work in the household, assistance to relatives, etc.). Finally, we estimate our outcomes at the metropolitan region level

Welfare Data: We use monthly averages by municipality on the share of workers who requested a dengue-related sick-leave for the years 2010-2017. This information is taken from the INSS, and dengue-specific leaves are identified using dengue's ICD-10 code (A90). Only workers who pay social security contributions can claim this benefit -only formal workers-, and only after the employer has paid 15 days of sick-leave.

There is no similar welfare scheme for informal workers. However, Brazil has one of the biggest conditional cash transfers in the world, Programa Bolsa Familia (PBF). Although this program is designed to attack structural poverty rather than temporary income shocks, it is possible that, if informal workers loose their job due to the disease, they may sign up to receive a cash transfer. We will use monthly averages by municipality on the number of PBF beneficiaries for the years 2011-2017.

Morbidity Data: We obtain our measures of morbidity from the National System of Information on Hospitalizations (SIH), a management system through which health facilities claim payments for all hospital admissions sponsored by the National Health Service from the Ministry of Health. The database contains information on all hospital episodes that took place within the public health care system in Brazil. We construct a panel at the municipality level of monthly hospitalization cases by diagnose for the period of January 2000 to April 2017.

Summary Statistics

Table B.3 provides descriptive statistics for the variables used in our labor market outcome analysis. Here one observation is one metropolitan-region/month combination. During the post-dengue period, on an average month a metropolitan-region has about 64% of the sample in the labor force, of which 92% are employed. Only 1% of those employed are on leave on an average week, while the rest work about 41 hours. Their earnings are 2,027$R (real Reais base Jan. 2016). In an average month, there would be 2,392 reported dengue cases. The values are similar for the pre-dengue period, except that on average there are less dengue reported cases (325) and a lower monthly income (1,655$R).

Table B.4 provides descriptive statistics for the variables used in our morbidity analysis. Here one observation is one municipality/month combination. As before, most values are similar between the two periods, except for those of dengue. An average municipality had 1.07(0.61) dengue hospi-

outbreaks than the other regions. However, when there is an outbreak in the central central region (specially Belo Horizonte and Rio de Janeiro) the incidence can be ten times higher than in the north. Finally, the only region in the south (Porto Alegre) shows a very small incidence of dengue, although it seems to be on the rise. Hence most of the effect we will estimate will come from the dengue outbreaks in the central region. Figures B.2 and B.2 repeat this for out weather variables: temperature and relative-humidity.

talizations during the post(pre) dengue period, and about 12.06(3.91) GP dengue cases. Average DSI is slightly higher in the pre-dengue period.

Finally, Table B.5 provides descriptive statistics for the welfare variables. The values are computed for an average municipality/month combination for the period 2011-2016 -no pre-dengue data available. An average municipality has on average 2,476 families with Bolsa Familia, which adds up to about 380,663$R per month. Similar to the leave cases in the PME sample, there are very few people taking dengue-related sickness-leave (1.3%)

3.2.3 Results

3.2.3.1 Labor Market Outcomes

Each cell in Table 3 reports the coefficient on (standardized) notified dengue cases (D) from Equation (1). The first four columns report the OLS estimates, while the last four columns report 2SLS estimates using DSI as instrument. Within estimation methods, each column uses a different set of time controls. Column 1(5) includes only year FE. Column 2(6) includes year plus month FE. Column 3(7) includes year times month FE. Model 4(8) includes year plus month FE but adds region-specific time-trend. Each row reports coefficient on a different outcome.

Our OLS estimates in column 1 are statistically significant and of the expected sign, with dengue cases having a negative effect on labor market outcomes. However, after adding different controls, the coefficient that is consistent in size and statistically significance is the share of people on sick-leave. Point estimates are qualitatively similar between OLS and 2SLS estimates (column 4 and 8), but 2SLS tend to be larger in magnitude. In the 2SLS estimates we find that dengue has a negative and statistically significant effect on hours and the share of people on sick-leave across all specifications. The coefficient on (log) earnings is less consistent, showing a positive and statistically significant coefficient when month FE are not included, and a negative coefficient on the other models. Results on labor force participation and employment are in general not statistically significant.

Our prefer specification includes an additive control for month and year FE, and region-specific time-trends (column 4). Using estimates from this model, a 1 standard deviation (sd) change in the number of dengue cases reduces hours worked in the previous week by 3.2% (about 5.2 hours a month) for workers with positive hours. This is a loss in productivity of about 60$R a month -5.2 hours times average hourly wage in our data (11.5$R). On the extensive margin, a one sd in dengue cases rises the share of people on sick-leave the previous week by 3.5 p.p. -a rise of 233% over the mean of 1.5%.

Heterogeneity analysis

Table 3: Effect of dengue on labor market outcomes

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) 2SLS	(7) 2SLS	(8) 2SLS
LFP	-0.004***	-0.001	-0.001	-0.000	-0.010***	-0.002	-0.003	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.009)	(0.008)	(0.009)
Emp.	-0.005***	-0.001	-0.001	-0.002**	0.001	-0.013	-0.007	-0.013*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.009)	(0.007)	(0.008)
Leave	0.001*	0.001**	0.001*	0.002***	0.019***	0.035***	0.030***	0.035***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)	(0.010)	(0.007)	(0.010)
Hours	-0.004***	-0.002	-0.002	-0.002	-0.029***	-0.029*	-0.019*	-0.032**
	(0.001)	(0.002)	(0.001)	(0.002)	(0.005)	(0.015)	(0.011)	(0.015)
Earnings	-0.035***	-0.002	-0.004	0.004	0.059***	-0.071*	-0.039	-0.062*
	(0.005)	(0.004)	(0.004)	(0.003)	(0.021)	(0.036)	(0.028)	(0.035)
Obs.	642	642	642	642	642	642	642	642
Year FE	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Month FE	No	Yes	No	Yes	No	Yes	No	Yes
Time FE	No	No	Yes	No	No	No	Yes	No
Region t-trend	No	No	No	Yes	No	No	No	Yes
IV F-stat					91	14	21	14

Notes: Each coefficient represents a separate regression of a labor market outcome on the number of GP dengue reported cases. The sample includes all 6 metropolitan-regions represented in PME and regional averages are calculated for people between 15 and 65 years old, excluding public employees, employers, and people with no remuneration. Outcome notation: LFP is the population share in the labor force; Emp is the share of the LF that are employed; $Leave$ is the share of employees who are absent the week before; $Hours$ is the logarithm of weekly hours worked; $Earnings$ is the logarithm of monthly earnings. Dengue cases (independent variable) is standardized as a Z-score. Coefficients are interpreted as percentage points (or % for hours and earnings) change in the outcome variable associated with a 1 sd increase in $Dengue$. 2SLS estimates use DSI (Z-score) as instrumental variable. All models include metropolitan-region fixed effects. Robust standard errors in parentheses, significance ***p<0.01, **p<0.05, *p<0.1

We first look at heterogeneous effects by type of job, diving the working population between those in formal jobs, informal jobs, and self-employed workers. Because we use aggregate data on dengue counts, our estimates of the effect of dengue on each group will confound to effects. On one side, the coefficient will be larger if group g has a bigger share in the number of dengue cases (more vulnerable to the disease). Secondly, the coefficient would also be larger if group g is more vulnerable (lower legal protection).

To try and disentangle these two channels, we resort to previous epidemiological literature on dengue and its socio-economic (SE) and demographic determinants. Teixeira et al. (2013) review the existing literature for Brazil and find that Afro-Brazilian/African ethnicity individuals are less likely to get dengue than whites (a factor of about 1 to 4) and males are slightly less likely to get it than women. Although the evidence on SE status and dengue incidence is more mixed, there seems to be the case that lower educated/poorer individuals are more likely to get dengue.

We show the share of males, whites and individuals with atleast secondary education for each group in Table 4. Men and white individuals are more or less equally represented in the three groups, but there is a higher proportion of highly-educated individuals in the formal group (62%) than in the other two (46% and 40% for informal and self-employed workers respectively). Given this, observe differences in coefficients would mainly be due to economic vulnerability between informal and self-employed workers, while formal workers benefit from lower vulnerability both economically and towards the disease.

Table 4: Descriptive statistics by sector

	Formal (58%)		Informal (16%)		Self-empl. (26%)	
	mean	sd	mean	sd	mean	sd
Hours worked (p/week)	43	1	39	2	40	2
Earnings (monthly)	1,817	373	1,248	345	1,559	438
% male	0.61	0.02	0.61	0.03	0.61	0.03
% white	0.49	0.23	0.47	0.24	0.48	0.25
% high education	0.62	0.08	0.46	0.06	0.40	0.06

Notes: Authors calculation from PME 2008-2014. The sample includes all 6 metropolitan-regions represented in PME and regional averages are calculated for people between 15 and 65 years old, excluding public employees, employers, and people with no remuneration.

Table 5 presents the 2SLS estimates of the effect of dengue cases on the share of workers on sick-leave, the number of hours worked and earnings by type of occupation. We can see that self-employed workers are the ones reducing the most their hours worked when dengue incidence is higher. They also observe a statistically significant drop in total earnings. As expected, formal workers are more protected and do not observe a statistically significant drop in hours nor earnings. Informal workers are somewhere in the middle.

However, all three types of workers have a very similar effect on the share of individuals on sick-leave -they all experience a rise of about 4.3 p.p.- although their group-means are different. This is consistent with the idea that, independently of the sector, a very sick worker will need to take time away from work independently of the sector. But those who keep working do seem to adjust differently in the intensive margin -number of hours work- by sector.

Table 6 repeats the exercise now breaking down the data by gender. We observe that all three coefficients on sick-leave, hours and earnings are larger for female workers. This may be due to the fact that women get more dengue, or that women are also more likely than man to loose hours of work to take care of a another household member. However, we can not reject the hypothesis that the gender differences are statistically equal to zero.

Robustness checks

The greatest threat to the validity of our analysis is that DSI may pick up climate factors that affect workers productivity directly (see literature review in Section 1). To lower this concern, we will use a placebo period -a period with none or very low dengue- but similar weather conditions. As mentioned before, Figure 3 shows that between 2003 and 2007, there were very few dengue cases. But starting 2008, almost every year observed a dengue outbreak.

Panel A in Table 7 shows the reduced-form effect of DSI on labor market outcomes using the same sample as our instrumental variable estimates of the effect of dengue before. Results are qualitatively the same. Panel B expands the sample to all years available (2003/2014) and adds an interaction term between DSI and a dummy for post-dengue years (after 2007). We expect the coefficient of this interaction to be statistically significant if DSI has a different effect in pre and post-dengue years. Indeed, this is what we find. A one sd increase in DSI lowers the working hours by 0.3% and earnings by 0.6% during the post-dengue period (with no effect during the pre-dengue period). DSI has a direct effect on the probability of being on leave during the pre-dengue period, and the effect in post-dengue years is larger and statistically different. This implies that our instrumental variable estimates oof the effect of dengue on the probability of being on-leave may be over estimated due to the direct effect of DSI (weather) on this outcome.[13]

3.2.3.2 Welfare Outcomes

This section looks at the relationship between dengue cases and two welfare schemes: (i) dengue-related paid-leaves, and (ii) Bolsa Familia beneficiaries. The first can only be used by formal workers while the second one is targeted at poor families (most likely informal workers). However, another difference is that the first one is designed to help during transitory health shocks, while the second one is designed to solve structural poverty issues (not transitory shocks). Given these difference in

[13] Tables B.1 and B.2 in Appendix B.1 repeat this exercise dividing workers by type and gender, and results are qualitatively the same as the instrumental variable estimates.

Table 5: Effect on hours worked and earnings: by type of occupation

	Formal			Informal			Self-employed		
	Leave	Hours	Earn.	Leave	Hours	Earn.	Leave	Hours	Earn.
Dengue	0.048***	-0.019	-0.042	0.037***	-0.036*	-0.060	0.043***	-0.065***	-0.109*
	(0.014)	(0.013)	(0.034)	(0.011)	(0.019)	(0.055)	(0.013)	(0.023)	(0.056)
Obs.	648	648	642	648	648	642	648	648	642
IV F-stat	14	14	14	14	14	14	14	14	14
Mean dep.	0.024	3.750	7.522	0.009	3.662	7.160	0.012	3.688	7.372

Notes: Each coefficient represents a separate regression of a labor market outcome on the number of GP dengue reported cases. The sample includes all 6 metropolitan-regions represented in PME and regional averages are calculated for people between 15 and 65 years old, excluding public employees, employers, and people with no remuneration. Outcome notation: *Leave* is the share of employees who are absent the week before; *Hours* is the logarithm of weekly hours worked; *Earnings* is the logarithm of monthly earnings. Dengue cases (independent variable) is standardized as a Z-score. Coefficients are interpreted as percentage points (or % for hours and earnings) change in the outcome variable associated with a 1 sd increase in *Dengue*. 2SLS estimates use DSI (Z-score) as instrumental variable. All models include fixed effects for metropolitan-regions, years, and month, and a metropolitan-region specific time-trend. Robust standard errors in parentheses, significance ***p<0.01, **p<0.05, *p<0.1

Table 6: Effect on hours worked and earnings: by gender

	Male			Female		
	Leave	Hours	Earn.	Leave	Hours	Earn.
Dengue	0.032***	-0.029**	-0.042	0.037***	-0.037**	-0.095**
	(0.009)	(0.014)	(0.036)	(0.011)	(0.016)	(0.042)
Obs.	648	648	642	648	648	642
IV F-stat	14	14	14	14	14	14
Mean dep.	0.0150	3.752	7.633	0.0150	3.661	7.463

Notes: Each coefficient represents a separate regression of a labor market outcome on the number of GP dengue reported cases. The sample includes all 6 metropolitan-regions represented in PME and regional averages are calculated for people between 15 and 65 years old, excluding public employees, employers, and people with no remuneration. Outcome notation: *Leave* is the share of employees who are absent the week before; *Hours* is the logarithm of weekly hours worked; *Earnings* is the logarithm of monthly earnings. Dengue cases (independent variable) is standardized as a Z-score. Coefficients are interpreted as percentage points (or % for hours and earnings) change in the outcome variable associated with a 1 sd increase in *Dengue*. 2SLS estimates use DSI (Z-score) as instrumental variable. All models include fixed effects for metropolitan-regions, years, and month, and a metropolitan-region specific time-trend. Robust standard errors in parentheses, significance ***p<0.01, **p<0.05, *p<0.1

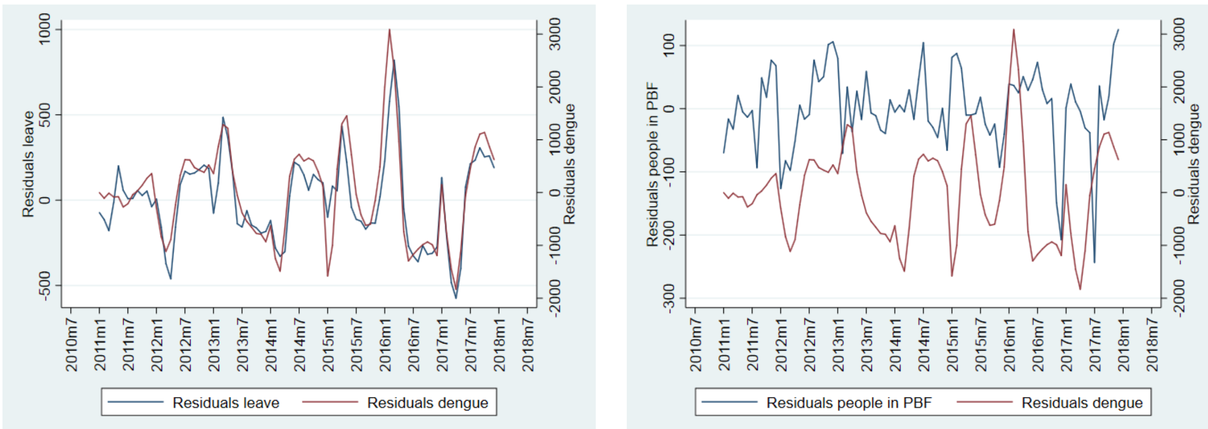
Table 7: Effect of dengue in the pre and post-dengue periods

	(1) LFP	(2) Emp	(3) Onleave	(4) Hours	(5) Earnings
Panel A: 2008/2014 sample					
DSI	-0.000	-0.001*	0.004***	-0.003**	-0.006**
	(0.001)	(0.001)	(0.000)	(0.001)	(0.003)
Panel B: 2003/2014 sample					
DSI	-0.001	-0.000	0.002***	-0.001	0.002
	(0.001)	(0.001)	(0.000)	(0.001)	(0.003)
DSI x Post	0.000	-0.001	0.001***	-0.003***	-0.006**
	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)

Notes: In Panel A, each coefficient represents a separate regression of a labor market outcome on the Dengue Suitability Index (DSI) for the period 2008/2014. In Panel B, there are two coefficients estimated: the first on is the effect of DSI on outcome during the 2003/2007 period; the second coefficient is added effect of DSI on outcome during 2008/2014. The sample includes all 6 metropolitan-regions represented in PME and regional averages are calculated for people between 15 and 65 years old, excluding public employees, employers, and people with no remuneration. Outcome notation: LFP is the population share in the labor force; *Emp* is the share of the LF that are employed; *Leave* is the share of employees who are absent the week before; $Hours$ is the logarithm of weekly hours worked; $Earnings$ is the logarithm of monthly earnings. DSI (independent variable) is standardized as a Z-score. Coefficients are interpreted as percentage points (or % for hours and earnings) change in the outcome variable associated with a 1 sd increase in *Dengue*. All models include fixed effects for metropolitan-regions, years, and month, and a metropolitan-region specific time-trend. Robust standard errors in parentheses, significance ***p<0.01, **p<0.05, *p<0.1

design, we expect paid-leaves to closely follow dengue outbreaks but not PBF beneficiaries. Figure 4 confirms this. We obtain the residual variation in dengue, paid-leaves and PBF beneficiaries after controlling for year plus month FE, municipality FE, and municipality-specific time-trend. On the left side figure (4a) we can see that the residuals of dengue and paid-leaves track each other very well in time. This is not true for the residuals of dengue and PBF beneficiaries (4b).

Figure 4: Residuals for dengue and welfare schemes



(a) Dengue and SS's dengue-related leaves

(b) Dengue and PBF beneficiaries

3.2.3.3 Health outcomes

One concern that may arise with our instrument is that weather variations may be affecting people's health trough a channel other than dengue. We take care of this by repeating our analysis using as outcome diagnose-related hospitalization counts (Table 8). Column (1) presents results for dengue-related hospitalizations. Columns (2) to (4) present results by type of episode (infectious disease, chronic and external). Columns (5) to (8) present results for non-communicable diseases: neoplasms (e.g. cancer tumour), diabetes, cardio, respiratory. Panel A shows second stage coefficients using DSI as instrumental variable for dengue. Panel B shows reduced-form estimates of DSI and its interaction with a dummy for post-dengue years (2008-2014). For comparability, all outcomes have been standardized to have a mean zero and a standard deviation of one. Because we have a large dataset (over 5,000 municipalities and 10 years of monthly data), we use an adjusted F critical value ($ln(n)$) to correct for large-sample over-rejection of the null hypothesis (first proposed by Leamer (1978) and Schwarz et al. (1978), and later popularized by Deaton (1997)).

In Panel A we observe that dengue-related hospitalizations show the largest correlation with dengue-reported cases. All but four of the F-statistics are significant at conventional significance levels. However, once we use an adjusted-large-sample F critical value ($ln(n) = 13.56$), the F-tests are smaller than this criterion for all types of hospitalizations but dengue-related ones. The next coefficient in size is that of infectious diseases hospitalizations, which include dengue-related diagnoses.[14] The results in Panel B are very similar qualitatively. The effect of DSI is only statistically significant on dengue hospitalizations. The fact that none of the other type of episodes are significantly correlated with GP dengue reported makes us more confident that we our estimation strategy is in fact capturing the effect of dengue on outcomes.

[14]Although not significant, the coefficient for chronic-disease hospitalizations is the third largest. One possible explanation is that some asthma cases may be coded as dengue by the general practitioner and reported as such. This would create a spurious positive correlation between notified dengue cases and asthma hospitalizations.

Table 8: Effect on hospitalizations by diagnose

	Dengue	Infectious	Chronic	External	Non-comunicable diseases			
					Neoplasms	Diabetes	Cardio	Respiratory
Panel A: IV estimates								
Dengue (GP)	1.0364	0.4057	0.1461	-0.1652	0.0003	0.0468	-0.0602	0.1128
	(0.1807)	(0.1139)	(0.0573)	(0.0671)	(0.0340)	(0.0357)	(0.0565)	(0.0879)
F-test	32.90	12.68	6.510	6.060	0	1.710	1.140	1.650
Panel B: RF estimates								
DSI	0.0355	0.0183	0.0088	-0.0088	0.0007	0.0048	-0.0043	0.0043
	(0.0129)	(0.0126)	(0.0044)	(0.0027)	(0.0026)	(0.0028)	(0.0036)	(0.0076)
DSI x Post	0.0690	0.0181	0.0022	-0.0048	-0.0013	-0.0033	0.0004	0.0067
	(0.0125)	(0.0060)	(0.0027)	(0.0028)	(0.0024)	(0.0031)	(0.0020)	(0.0046)
F-test (inter.)	30.38	9.170	0.630	2.900	0.280	1.150	0.0400	2.130
Observations	771,684	771,684	771,684	771,684	771,684	771,684	771,684	771,684

Notes: In Panel A, each coefficient represents a separate regression of the number of diagnose-specific hospitalization on the GP dengue reported cases for the period 2003-2014. In Panel B, there are two coefficients estimated: the first on is the effect of DSI on outcome during the 2003/2007 period; the second coefficient is the added effect of DSI on outcome during 2008/2014. The sample includes 5,359 municipalities. Outcomes and independent variables (dengue cases and DSI) are standardized as Z-scores, so coefficients are interpreted as stadard deviations (sd) change in the outcome variable associated with a 1 sd increase in the independent variable. All models include fixed effects for metropolitan-regions, years, and month, and a metropolitan-region specific time-trend. The adjusted F-test critical value is ($ln(n) = 13.56$). Standard errors clustered at state level in parentheses.

3.3 Micro Effects

3.3.1 Data and Empirical Framework

We now investigate dengue effects on outcomes at the household and individual levels, by combining identified administrative data on both dengue infections and family records on socioeconomic outcomes and access to welfare and social assistance. In this section we present the data and the empirical model.

3.3.1.1 Data

Dengue Data: We use identified data from all the notified dengue cases during the dengue season of 2017 reported in SINAN by the GPs. We describe this data in at length in Appendix A.2.

Welfare data: We use identified data from all people registered in CadUnico, a national registry system that aims to map over time all families and individuals of low socio-economic status in the country. We describe this data in at length in Appendix A.1.

3.3.1.2 Sample and Final Dataset

On CadUnico, we have the set of families/individuals who updated records up to April 2017. Therefore, we can observe socioeconomic outcomes for families who may have updated their records in CadUnico before or after contracting dengue.

From CadUnico, we first select families who updated their records both in 2016 and 2017, and then draw all individuals that belong to those families. All individuals on SINAN are included in the sample matched with CadUnico. There are duplicated entries of the same case in different facilities but we keep the entry registered in the earliest date of appearance of symptoms. Records between SINAN and CadUnico are then matched using individuals names, exact date of birth, gender and municipality of residence. The matching between names are evaluated based on Levenshtein distances, those above 0.88 are taken as true matches. This threshold was determined by an eyeballed random sample of names.

In 2017, there were approximately 26.7 million families on CadUnico, and 88 million individuals. SINAN had 207,239 dengue cases reported during the 2017 season. Given the sub-sample of families that updated their records both in 2016 and 2017, up to April, and restricted to municipalities of residence with at least one reported case of dengue, we end up with a sample of 1,888,209 different individuals, 571,461 families, located in 925 different municipalities – predominantly in the Northeast and Southeast regions, the ones with higher dengue incidence. Out of the final sample, we were able to find 7,609 individuals with a reported case of dengue.

Table C.1 presents descriptive statistics based on data from the baseline year, 2016. We compare baseline socioeconomic indicators for households with *vs* without any individuals that contracted dengue during the 2017 dengue season. In the upper panel, we restrict the sample to households located in municipalities with at least one reported case of dengue in 2017. Next, we restrict the sample to households located in zipcodes with at least one reported case of dengue; in the bottom panel, we further restrict the sample to those households with children enrolled in schools for which we identify children or any of their respective family members that contracted dengue in the 2017 season.

Overall, we observe that statistics for households without any dengue case change across panels, and that differences between groups within panels are statistically different. In particular, we observe better socioeconomic indicators for those households that experienced a case of dengue in 2017, which is consistent with the fact that dengue incidence is typically higher in urban and metropolitan areas. As discussed in the next section, we exploit different samples, the timing of the disease, and rely on a difference-in-differences (DiD) specification in order to balance treatment *vs* control groups, so to recover a causal interpretation from our estimates.

3.3.1.3 Estimating Equation

We rely on a difference-in-differences specification to examine the extent to which dengue fever contraction by any family members impacts family income per capita. In particular, we focus on the head of household. Exposure to dengue at the very local level can be considered exogenous, although susceptibility to contraction and infection severity typically depend on previous health conditions, thus correlated with living conditions and SES. In our empirical strategy we exploit a series of local and/or family fixed-effects and the timing of the first symptoms in order to identify effects as well as to provide robustness checks that are typical to DiD strategies. The following equation provides our conceptual setup:

$$Sine(y_{imt}) = \alpha_i + \phi_{mt} + \beta Dengue_{imt} + X'_{it}\Gamma + u_{imt}$$

Where $Sine(y_{imt})$ is the inverse hyperbolic sine transformation of the monthly family income per capita for individual i, recorded for month t, located in municipality m. Family income typically includes labor market earnings of all family members as well as the sum of social security pensions, donation, alimony, CCT and other welfare transfers. In our benchmark specification, i refers to the head of household, typically female in CadUnico records. The term α_i refers to individual fixed-effects, while ϕ_{mt} corresponds to municipality-year-month fixed-effects – more specifically, we include a dummy for each combination of municipality, year and month of update in CadUnico. This set of fixed-effects absorbs the influence of the length time in between updates. Our variable of interest is $Dengue_{imt}$, an indicator that identifies whether individual i (the head of household or other family member) contracted dengue in month t during the 2017 dengue season, up to April. Because we have the exact date of the first symptoms, we are also able to use more flexible ways of computing dengue shocks – for instance, by the number of days or weeks before and after the day of the first symptoms. Finally, the term X_{it} includes a number of controls for household and individual's time-varying characteristics, such as (i) dummies for access to potable water, sanitation, garbage collection, electricity and road pavement; and (ii) school attainment, age and square age, and an indicator for agricultural employment.

We provide support for the parallel pre-trends assumption by testing for the timing of the first symptoms, before *vs* after CadUnico update. A relevant caveat otherwise refers to the fact that we do not observe dengue infections in the 2016 season, so our baseline is identical across individuals. We overcome this issue by including individual fixed-effects, thus absorbing initial conditions. Furthermore, we also restrict the sample to those families for which the distance between updates is equal or less than 12 months, thus relying on a sub-sample of individuals whose baseline barely includes the 2016 season. In this case, we minimize the likelihood of assigning non-exposure to dengue for those that actually may have contracted the disease in the baseline year.

3.3.2 Results

3.3.2.1 Dengue Effects on Monthly Family Income

The main results are reported on Table 9. Panel A presents a series of specifications based on our benchmark equation. Samples across all regressions are restricted to the head of household, and our variable of interest identifies whether the head of household was hit by dengue in the month just before the CadUnico update, for which we have the family income per capita in that month. In the first column we include only municipality-year-month fixed-effects. We then progressively add controls for household characteristics (column 2) and for individual characteristics (column 3). In the first three columns we restrict the sample to households located in municipalities with at least one reported case of dengue in 2017. In column 4 we restrict the sample to households located in zipcodes with at least one reported case of dengue, while in column 5 we further restrict the sample to those households with children enrolled in schools for which we identify children or any of their respective family members who contracted dengue in the 2017 season. Finally, in the remainder column we restrict the sample only to families that had a case of dengue during the 2017 season – not necessarily before the CadUnico update. More specifically, in this latter sample, we compare all families that were hit by dengue, but at different points in time, before *vs* after the update.

Overall, we observe a negative, robust and remarkably stable point estimate across the first 5 columns, irrespectively of sample selection and controls being considered. We observe that once the head of household is hit by dengue, the family income per capita drops by approximately

28% in that given month. In column 6 the point estimate increases twofold, which suggests that measurement error in dengue detection might play an important role. In that sense, the estimated effect of 28%, although relevant, can be still considered a lower bound.

In Panel B we test an alternative and more demanding specification, in which the dependent variable is computed as the change in the monthly family income per capita between updates. In that case, conditioning upon controls and fixed-effects corresponds to controlling for specific time trends on characteristics. We find very similar and still significant point estimates (at 10%) across columns. In column 6 the coefficient increases by half, it is no longer significant, but supports a qualitatively identical interpretation.

Table 9: Dengue effect on monthly family income

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A - DiD - Dep Var: Sine(Family Income per Capita)						
Dengue Month Before - Head of HH	-0.253	-0.277	-0.281	-0.284	-0.298	-0.586
	(0.124)**	(0.131)**	(0.135)**	(0.142)**	(0.138)**	(0.280)**
Observations	980,548	864,372	820,268	489,224	373,588	2,422
R-squared	0.825	0.819	0.819	0.864	0.813	0.863
Panel B - FD on Level - Dep Var: Delta [Sine(Family Income per Capita)]						
Dengue Month Before - Head of HH	-0.231	-0.244	-0.231	-0.248	-0.224	-0.341
	(0.123)*	(0.129)*	(0.131)*	(0.136)*	(0.132)*	(0.211)
Observations	494,825	443,106	423,205	258,720	194,806	2,018
R-squared	0.066	0.069	0.072	0.105	0.079	0.259
MonthUptade*YearUpdate*Municip	Yes	Yes	Yes	Yes	Yes	Yes
HH Controls	No	Yes	Yes	Yes	Yes	Yes
Indiv Controls	No	No	Yes	Yes	Yes	Yes
At least 1 case of dengue in :	Mun	Mun	Mun	Zipcode	School	Family

Notes: In all specifications the dependent variable is the inverse hyperbolic sine transformation of the monthly family income per capita for individual (Panel A) and its change between CadUnico updates (Panel B). The variable of interest is a dummy that identifies whether the individual (the head of household) contracted dengue in month before update, during the 2017 dengue season. Household controls include dummies for access to potable water, sanitation, garbage collection, electricity and road pavement; individual's controls include school attainment, age and square age, and an indicator for agricultural employment. Standard errors clustered at theindividual level, significance ***p<0.01, **p<0.05, *p<0.1

3.3.2.2 Dengue Effects by Family Member

Next, we examine whether the effects on family income depend on the member of the household that is actually hit by dengue – we compute the indicator of dengue in the month before the update for the different members of the household, more specifically, for the head of household, her/his respective spouse, and the children. It is important to note that about 93% of the heads of household in our sample correspond to female individuals, which is consistent with CadUnico records being generally filled in by women – the ones who are typically the recipients of social benefits and cash transfers. In that sense, the results from Table 9 suggests that households become particularly vulnerable once women are hit by dengue. This is further confirmed on Table 10. We replicate the same specifications from columns 3-6 of Table 9, and examine the shock by household members. We find that effects on family income are particularly relevant when it is the head of household who is hit by the infection. We observe insignificant coefficients for the spouse as well as for the children, except in column 4. In that latter specification, restricted to families with at least one case of dengue during the 2017 season, we find, again, greater effects for the head of household, and a negative and robust effect for the children. As child labor is limited in our context, this finding suggests that women may reallocate time from the labor market to the household in order to care for ill children.

Table 10: Dengue effect by family member

	Sine(Family Income per Capita)			
	(1)	(2)	(3)	(4)
Dengue Month Before - Head of HH	-0.283	-0.286	-0.298	-0.643
	(0.135)**	(0.142)**	(0.138)**	(0.278)**
Dengue Month Before - Spouse	0.233	0.133	0.273	-0.016
	(0.167)	(0.164)	(0.182)	(0.306)
Dengue Month Before - Children	-0.030	0.024	-0.092	-0.402
	(0.086)	(0.085)	(0.089)	(0.164)**
Observations	820,268	489,224	373,588	2,422
R-squared	0.819	0.864	0.813	0.864
MonthUptade*YearUpdate*Municip	Yes	Yes	Yes	Yes
TimesinceUpdate FE	Yes	Yes	Yes	Yes
HH Controls	Yes	Yes	Yes	Yes
Indiv Controls	Yes	Yes	Yes	Yes
Sample Geo restricted to:	Mun	Zip	School	Family

Notes: In all specifications the dependent variable is the inverse hyperbolic sine transformation of the monthly family income per capita for individual. The variable of interest is a dummy that identifies whether the individual (the head of household, her spouse or children) contracted dengue in month before update, during the 2017 dengue season. Household controls include dummies for access to potable water, sanitation, garbage collection, electricity and road pavement; individual's controls include school attainment, age and square age, and an indicator for agricultural employment. Standard errors clustered at theindividual level, significance ***p<0.01, **p<0.05, *p<0.1

3.3.2.3 Dengue Effects: Timing

We now break the computation of the dengue indicator by 10-day bins, before and after the CadUnico update, in order to examine the timing of the effects. We identify the exact day of the first symptoms, and define time dummies accordingly, focusing on the head of household. More specifically, we compute an indicator for whether the head of household contracted dengue within the ten days just before the update; within the 10-20 days before; 30-40 days before; as well as within the ten days just after the update. We follow the same series of specifications of Table 10, and examine both the DiD and the first-difference on levels, as in Table 9.

Table 11 presents the results. Overall, the strongest effects come from the shocks by the beginning of the month, between 20-30 days before the update. On Panel A, we also find smaller but still robust and non-trivial impacts for the 10-20 bin. These results are consistent in light of the fact that family income is computed over earnings within the 30-day period before the update, and that dengue infections typically last for 7-10 days from the day of the first symptoms. Once the individual is hit by the beginning of the month, most of her time and work capacity within the month is therefore consumed by the illness.

Table 11 also indicates that dengue effects are transitory, as coefficients for dengue shocks in the month before the update are insignificant and smaller in magnitude (for the 30-40 bin before update). We also find insignificant coefficients for the 10-day period just after the update, which works as a placebo exercise and reassures the validity of our empirical strategy.

3.3.2.4 Dengue Effects: Welfare Dependency and Mitigation Mechanism

We now examine the extent to which dengue shocks may trigger welfare dependency, and whether welfare programs, such as PBF, may act as an insurance scheme against illness. The first four columns of Table 12 follow the same sequence of specifications from Table 10, but also include an interaction term between dengue and enrollment in PBF. In the remainder columns, the dependent variable is a dummy for enrollment in PBF. In this case, we test whether a transitory health shock is associated with an increase in access to welfare as measured by enrollment in a CCT program.

The first four columns show that PBF mitigates nearly all detrimental effects of dengue on family income, and can be interpreted as an insurance against a transitory health shock. On the other hand, in the remaining columns we observe positive but insignificant effects of dengue shocks on PBF enrollment. This means that, despite a CCT program such as PBF being effective as an insurance scheme in times of vulnerability, in particular for the poorest and the informal households, as those registered in CadUnico, access to transfers seems not to respond in a timely manner to household needs. The results of Table 12 combined thus suggest that transitory insurance schemes for the most vulnerable families may be effective in mitigating detrimental impacts of health shocks.

Table 11: Timing of the effect

	(1)	(2)	(3)	(4)
Panel A - DiD - Dep Var: Sine(Family Income per Capita)				
Dengue 10 Days Before - Head of HH	0.242	0.239	0.265	-0.258
	(0.206)	(0.225)	(0.218)	(0.424)
Dengue 10to20 Days Before - Head of HH	-0.425	-0.374	-0.470	-0.948
	(0.251)*	(0.264)	(0.261)*	(0.643)
Dengue 20to30 Days Before - Head of HH	-0.652	-0.671	-0.668	-0.560
	(0.229)***	(0.231)***	(0.227)***	(0.352)
Dengue 30to40 Days Before - Head of HH	0.029	0.021	-0.019	-0.271
	(0.198)	(0.207)	(0.191)	(0.351)
Dengue 10 Days After - Head of HH	0.267	0.202	0.305	0.027
	(0.234)	(0.232)	(0.238)	(0.322)
Observations	820,268	489,224	373,588	2,422
R-squared	0.819	0.864	0.813	0.863
Panel B - FD on Level - Dep Var: Delta [Sine(Family Income per Capita)]				
Dengue 10 Days Before - Head of HH	0.261	0.272	0.308	0.489
	(0.190)	(0.201)	(0.199)	(0.352)
Dengue 10to20 Days Before - Head of HH	-0.363	-0.360	-0.409	-0.849
	(0.255)	(0.262)	(0.255)	(0.402)**
Dengue 20to30 Days Before - Head of HH	-0.612	-0.653	-0.601	-0.591
	(0.225)***	(0.228)***	(0.218)***	(0.285)**
Dengue 30to40 Days Before - Head of HH	0.062	0.036	0.140	-0.091
	(0.200)	(0.201)	(0.206)	(0.247)
Dengue 10 Days After - Head of HH	0.256	0.268	0.295	0.324
	(0.226)	(0.228)	(0.226)	(0.302)
Observations	423,205	258,720	194,806	2,018
R-squared	0.072	0.105	0.079	0.266
MonthUptade*YearUpdate*Municip	Yes	Yes	Yes	Yes
TimesinceUpdate FE	No	No	No	No
HH Controls	Yes	Yes	Yes	Yes
Indiv Controls	Yes	Yes	Yes	Yes
Sample Geo restricted to:	Mun	Zip	School	Family

Notes: In all specifications the dependent variable is the inverse hyperbolic sine transformation of the monthly family income per capita for individual (Panel A) and its change between CadUnico updates (Panel B). The variable of interest is a dummy that identifies whether the individual (the head of household) contracted dengue at different 10-day bins before and after update, during the 2017 dengue season. Household controls include dummies for access to potable water, sanitation, garbage collection, electricity and road pavement; individual's controls include school attainment, age and square age, and an indicator for agricultural employment. Standard errors clustered at the individual level, significance ***p<0.01, **p<0.05, *p<0.1

Table 12: Welfare dependency

	DiD - Dep Var: Sine(Family Income per Capita) and PBF							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep Var: Sine(Family Income per Capita)				**Dep Var: PBF**			
Dengue Month Before - Head of HH	-0.708	-0.781	-0.726	-1.117	0.025	0.025	0.024	0.021
	(0.297)**	(0.317)**	(0.300)**	(0.483)**	(0.019)	(0.020)	(0.019)	(0.034)
Dengue Month Before - Head of HH * PBF	0.613	0.690	0.620	0.821				
	(0.329)*	(0.349)**	(0.333)*	(0.547)				
Observations	820,268	489,224	373,588	2,422	820,268	489,224	373,588	2,422
R-squared	0.828	0.869	0.825	0.871	0.907	0.916	0.913	0.935
MonthUptade*YearUpdate*Municip	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TimesinceUpdate FE	Yes	Yes	Yes	Yes	No	No	No	No
HH Controls + PBF*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiv Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample Geo restricted to:	Mun	Zip	School	Family	Mun	Zip	School	Family

Notes: In all specifications in the first four columns the dependent variable is the inverse hyperbolic sine transformation of the monthly family income per capita for individual; in columns 5-8 is a dummy for PBF enrollment. The variable of interest is a dummy that identifies whether the individual (the head of household) contracted dengue in the month before the update, during the 2017 dengue season. Columns 1-4 include a dummy for PBF as control. Household controls include dummies for access to potable water, sanitation, garbage collection, electricity and road pavement; individual's controls include school attainment, age and square age, and an indicator for agricultural employment. Standard errors clustered at the individual level, significance ***p<0.01, **p<0.05, *p<0.1

4 A Positive Health Shock: Health-care facilities' opening

The Family Health Program (Programa de Saude da Familia - PSF) is a community-based approach to providing primary health care in Brazil. It relies on lay community health agents and interdisciplinary care teams to provide universal access to proactive first point of contact care. It includes general practice, community pharmacy, dental services and epidemiological surveillance.

Following guidelines from the Ministry of Health, the PSF was first implemented in small and medium cities throughout the 90's, as a smaller number of teams would be necessary to cover most of the population. In large cities it was implemented more incrementally, following difficulties in management and personnel. While nationally the PSF reached more than 90 percent of coverage by 2006, the municipality of Rio de Janeiro until December 2008 was the country's capital city with the lowest coverage of PSF teams: 3.5 percent of the population. However, starting from 2008 the program was extended largely and rapidly. Covering 3 percent of the city in 2008, it reached 52 percent in 2011 and 86 percent in 2017.

PSF has been consistently associated with improvements in child mortality and large reductions in maternal, foetal, neonatal and post-neonatal mortality (Bhalotra et al., 2019). It has also been shown to reduce hospitalizations by conditions sensitive to primary care (Macinko et al., 2010). Rocha and Soares (2010), in turn, find a positive effect on adult labor supply and employment that shows up in a municipality after eight years of PSF.

The late expansion in Rio – arguably exogenous – allows us to combine the implementation of PSF with CadUnico's (described in previous section and in Appendix A.1) rich longitudinal data that was collected mostly at the same time. These data allow us to pin down PSF causal effects at the family and individual levels with treatment evaluated at the catchment area of a health facility with PSF teams. By leveraging the impact of PSF on labor market outcomes we can demonstrate how health care policy with multiple benefits can further be turned into an unintended labor market policy.

4.1 Data

We build a novel dataset by combining administrative data from the Ministry of Health, the Ministry of Social Development and Rio de Janeiro's Secretariat of Health.

First, we geocode all CadUnico's household addresses – available from the Ministry of Social Development – using the Google Maps service. Addresses located in slum areas in Rio de Janeiro are particularly difficulty to identify. Precise data about them are not available from Google or any other geocoding platform. These locations are dominated by militias[15] or heavily armed drug

[15]Criminal groups made up of former soldiers, police officers and firemen, that use terror to control locals and

gangs that do not easily allow the entry of mapping services. To overcome these difficulties we have used maps available from Rio de Janeiro's Instituto Municipal de Urbanismo Pereira Passos (IPP) to identify those addresses not located by Google.

Next, we intersect the geographic coordinates of each address with information on the catchment areas of PSF teams based on a given health facility. These data are available from Rio's Secretariat of Health. If the address falls within a catchment area we consider it covered by PSF. However, we assume that coverage starts at the time the first PSF team starts working in the health facility, as registered in the CNES database available from the Ministry of Health.

PSF teams make community visits and ensure that all families in the catchment area are identified and visited regularly. This feature allows us to confidently rely on catchment areas for the allocation of treatment, differently from other types of medical services that do not have a well defined geographic coverage.

4.2 Econometric Specification

We start with a standard difference-in-differences specification that includes individual/family and time fixed effects:

$$y_{ifnt} = \alpha + \beta \times PSF_{nt} + X_{ifnt}\gamma + D_t + \epsilon_{ifnt} \tag{3}$$

where y_{ifnt} is the outcome of interest for individual i in family f. We keep the f subscript to indicate that some outcomes are observed at the family level only. PSF is a binary indicator that equals 1 when individual i lives within a catchment area n with PSF. D_t are month and year fixed effects. The parameter of interest β represents the difference-in-differences estimate of the labor market effect of PSF. Standard errors are either robust or clustered by administrative regions[16], depending on the specification.

Whenever appropriate, we also use the distributed lags specification below to examine the dynamic effects of PSF expansion. Because we observe labor market outcomes for families/individuals at two years intervals, approximately, lags above two years are identified only by families/individuals observed more than twice.

$$y_{int} = \alpha + \sum_{j=0}^{J} \beta_j \times PSF_{n(t-j)} + X_{ifnt}\gamma + D_t + \epsilon_{ifnt} \tag{4}$$

$CSF_{n(t-j)}$ is a dummy that assumes value 1 when the individual's address is covered by PSF in

businesses.

[16] Groups of neighborhoods under the same sub-council administration.

month/year $t - j$. Standard errors are also robust or clustered by administrative regions.

Primary health care is organized to focus on local problems and emphasizes preventive practice. For instance, the health problems faced by an area populated by industrial workers, such as the neighborhood of Campo Grande that houses an industrial district, is different from one populated by poor migrants, such as Rocinha. To account for the fact that this may generate dynamic variation that correlates with treatment we also include health facility-specific time trends in some specifications.

Identification relies on the assumption that the timing of PSF implementation is uncorrelated with other determinants of labor market outcomes at the catchment area level. Rio de Janeiro's Council had explicit guidelines to implement the program first in poorer areas, so the allocation of treatment is entirely based on initial conditions that are taken care of with fixed effects.

4.3 Sample Selection and Variable Definition

We select a sample of families for whom we observe at least two updates where one has been made before and another after the implementation of PSF in a given address within a catchment area. Therefore, the variation used for identification is in the timing of PSF implementation across neighborhoods, since all units in the analysis have eventually been treated.

Our sample is composed by 20,442 family-month-year of an unbalanced panel that has an average length of 18 months-time between one observation and the next. For individual earnings this sample goes up to 53,212 keeping the same restrictions.

Per capita family income is defined as the ratio between the total family income in the previous month, excluding any benefits, divided by the number of family members. At the individual level, labor earnings are those of the month prior to CadUnico's record update. `worked_prevweek` is 1 for those who worked in the previous week. In the baseline, those who missed work due to some illness are also included.

4.4 Results

We start by conducting an analysis at the family level because average family income per capita has fewer missing values and is expected to be less noisy for being an average. Table 13 shows the effects of PSF introduction on average family's per capita monthly income. We use the inverse hyperbolic sine transformation of income as outcome to treat zeros and large values. Column (1) displays the regression with month, year and family fixed effects. To increase statistical precision we also add two sets of controls: family and neighborhood characteristics. Family controls include indicators for households served with sewage, rubbish collection, electricity, paved roads, piped water and bathroom. Neighborhood controls include population size, dummies for other health care services available and transport infrastructure. Estimates point to a 17 percent increase in per capita family income as a result of the availability of PSF where they live. This number grows by two points when health facility specific fixed effects are added (Column (2)). This growth is statistically significant whether we add clustering by larger aggregation groups or not (Column (3)).

Table 13: PSF effects on per capita family income

	(1)	(2)	(3)
Dep.Var.:	sinh_income	sinh_income	sinh_income
PSF Catchment Area	0.176***	0.198***	0.198***
	(0.048)	(0.056)	(0.055)
Health Facility * year FEs		Yes	Yes
VCE	robust	robust	cluster
Cluster Var			Admin Region
Dep. Var. Sample Avg.	197.69	197.69	197.69
Observations	20442	20442	20442
R^2	0.65	0.67	0.67

Note: PSF Catchment Area == 1 for families living within the catchment area of a health facility with at least one PSF team. Standard errors are as indicated by the rows VCE and Cluster Var. Income is transformed using the inverse hyperbolic sine function to handle zeros and large values. All regressions include household and neighborhood controls, and fixed effects for month, year and family. Household controls include indicators for households with sewage and rubbish collection, electricity, paved roads, piped water and bathroom (all interacted with year). Neighborhood controls include population size and dummies for other health care services available and transport infrastructure. Standard errors in parenthesis, significance ***p<0.01, **p<0.05, *p<0.1

Table 14, in turn, splits these effects into intervals of two years since the implementation of PSF. In years 1 and 2, per capita family income raises by approximately 23 percent (Columns (2) and (3)), and in the following two years 26 percent. Coefficients from Column (2), our preferred specification, are depicted in Figure 5 for better visualization. Note that these coefficients are identified by different sets of families that potentially differ across important characteristics. In Table 15 we test for composition effects in these results. We run the same specifications as in the Table 13 but with a selected subsample where PSF should not have any effect on income because the included families had at least two updates both with PSF coverage. As it can be seen from specifications in Columns (2) and (3) the PSF effects for this sample is virtually zero.

In Table 16 we conduct the previous analysis at the individual level. Here, we include in our sample those who are representatives of the family at CadUnico's interview and their declared partners. Representatives are normally the mothers, a PBF requirement, and partners their husbands. At this level of analysis, we further include individual characteristics such as age, race and education

Table 14: Dynamics of the PSF effects on per capita family income

Dep.Var.:	(1) sinh_income	(2) sinh_income	(3) sinh_income
Pre-PSF Catchment Area 3+	−0.101	−0.101	−0.101
	(0.118)	(0.138)	(0.269)
Pre-PSF Catchment Area 1,2	−0.076	−0.076	−0.076
	(0.096)	(0.119)	(0.120)
PSF Catchment Area 1,2	0.169**	0.227**	0.227**
	(0.079)	(0.103)	(0.097)
PSF Catchment Area 3+	0.118	0.259*	0.259
	(0.099)	(0.134)	(0.246)
Health Facility * year FEs		Yes	Yes
VCE	robust	robust	cluster
Cluster Var			Admin Region
Dep. Var. Sample Avg.	197.69	197.69	197.69
Observations	20442	20442	20442
R^2	0.70	0.71	0.71

Note: PSF Catchment Area 1,2 (3+) == 1 (Pre-PSF Catchment Area 1,2 (3+) == 1) for families living within the catchment area of one PSF team that updated CadUnico records up to two years (three years or more) after (before) PSF arrival. Standard errors are as indicated by the rows VCE and Cluster Var. Income is transformed using the inverse hyperbolic sine function to handle zeros and large values. All regressions include household and neighborhood controls, and fixed effects for month, year and family. Household controls include indicators for households with sewage and rubbish collection, electricity, paved roads, piped water and bathroom (all interacted with year). Neighborhood controls include population size and dummies for other health care services available and transport infrastructure. Standard errors in parenthesis, significance ***p<0.01, **p<0.05, *p<0.1

interacted with time. In Column (1), we find an increase of approximately 13 percent in earnings that seem to be powered by a higher probability of being at work (2 percent, Column (2)). Splitting these results by gender, earnings go up by 16 percent for women, who were more likely to be working in the last week, whereas for men these effects are not statistically significant.

It is not difficult to find reasons why women would benefit more in the long term. For instance, these effects could partly be driven by a reduction in their role as caregivers for relatives with conditions sensitive to primary care. Considering the smaller sample of men, if anything, men also display gains in earnings, but less than half of that for women. Growth more likely to come from the intensive margin of work, as the probability of being at work is virtually zero. Something that could be explained by a reduction in the severity of conditions that are amenable to primary care, as well as preventable diseases, that affect men and women similarly. Put together, these results indicate that the PSF may have had an overall positive effect on labor market outcomes of CadUnico families.

Table 15: PSF effects on per capita family income for a subsample of families that always had PSF coverage

Dep.Var.:	(1) sinh_income	(2) sinh_income	(3) sinh_income
PSF Catchment Area	−0.055**	0.008	0.008
	(0.026)	(0.061)	(0.145)
Health Facility * year FEs		Yes	Yes
VCE	robust	robust	cluster
Cluster Var			Admin Region
Dep. Var. Sample Avg.	157.83	157.83	157.83
Observations	255099	255099	255099
R^2	0.65	0.66	0.66

Note: PSF Catchment Area == 1 for families living within the catchment area of a health facility with at least one PSF team. Standard errors are as indicated by the rows VCE and Cluster Var. Income is transformed using the inverse hyperbolic sine function to handle zeros and large values. All regressions include household and neighborhood controls, and fixed effects for month, year and family. Household controls include indicators for households with sewage and rubbish collection, electricity, paved roads, piped water and bathroom (all interacted with year). Neighborhood controls include population size and dummies for other health care services available and transport infrastructure. Standard errors in parenthesis, significance ***p<0.01, **p<0.05, *p<0.1

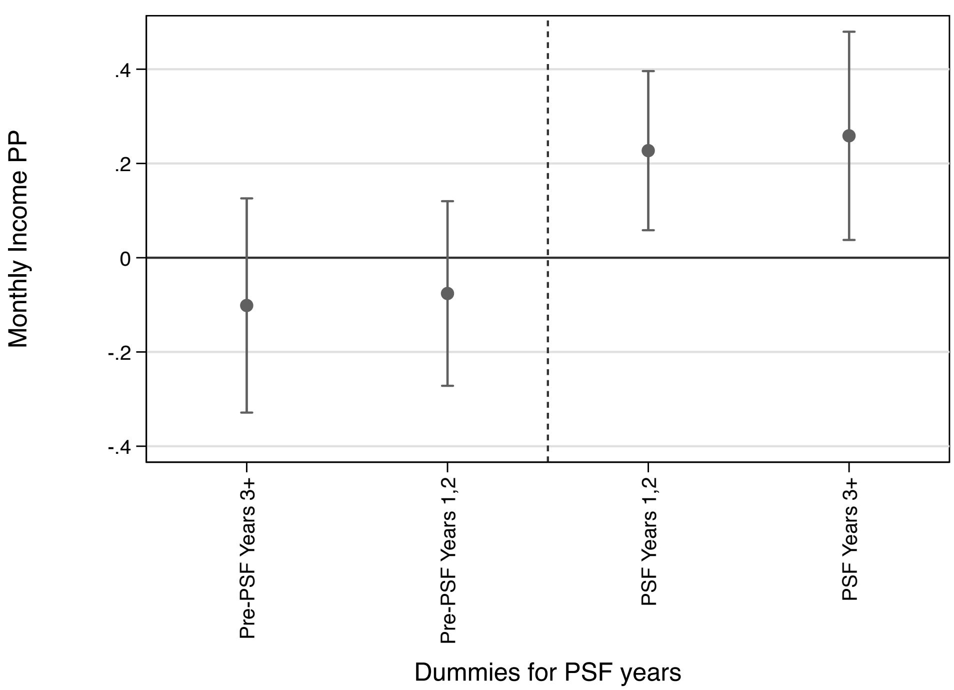


Figure 5: Dynamics of the PSF effects on per capita family income

Note: The graph displays coefficients for Table 14 Column(2).

Table 16: PSF effects on individual earnings

	(1) sinh_earnings All	(2) worked_prevweek All	(3) sinh_earnings Female	(4) worked_prevweek Female	(5) sinh_earnings Male	(6) worked_prevweek Male
PSF Catchment Area	0.135***	0.021***	0.157**	0.027***	0.062	0.002
	(0.050)	(0.008)	(0.063)	(0.009)	(0.085)	(0.012)
VCE	cluster	cluster	cluster	cluster	cluster	cluster
Cluster Var	Family ID	Family ID	Family ID	Family ID	Family ID	Family ID
Dep. Var. Sample Avg.	200.26	0.36	198.78	0.40	204.03	0.27
Observations	53212	53212	38171	38171	14979	14979
R^2	0.74	0.73	0.71	0.70	0.80	0.79

Note: PSF Catchment Area == 1 for families living within the catchment area of a health facility with at least one PSF team. Standard errors are as indicated by the rows VCE and Cluster Var. Income is transformed using the inverse hyperbolic sine function to handle zeros and large values. All regressions include individual, household and neighborhood controls, and fixed effects for month, year and individual. Household controls include indicators for households with sewage and rubbish collection, electricity, paved roads, piped water and bathroom (all interacted with year). Neighborhood controls include population size and dummies for other health care services available and transport infrastructure. Individual controls include age, race and education interacted with time. Standard errors in parenthesis, significance ***p<0.01, **p<0.05, *p<0.1

5 Final Remarks

In this paper we revisit an old question in the economics field: does health affect income and welfare? We answer this question using a combination of data and econometric models that allow us to obtain causal estimates and quantify these effects. We explore two forms of health shocks: first, we analyze the effects of a negative health shock (dengue outbreaks); second, we look at positive health shocks provided by the implementation and expansion of health-care programs (PSF).

We find that individuals and regions who are affected by dengue see their income and working hours substantially reduced. However, we also find that beneficiaries of Brazil's main conditional cash transfer program (PBF) do not see their family income drop. On the other hand, families who reside in the catchment area of a newly built primary care unit (PSF) observe an increase in their family per capita income. Given this evidence, we conclude that health is an important factor in determining individuals income and welfare dependency.

Our contribution to the literature is fourfold. First, we contribute with new causal estimates of the effects of health on income and welfare. Second, unlike most of the evidence in this literature, we use data for a developing country -Brazil. Third, we are the first to analyze the productivity effects of an infectious disease with the potential to become epidemic in a relatively short span of time. Finally, we also contribute to the literature looking at the effects of climate change on human productivity with a previously unexplored channel: infectious diseases.

References

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. *The Quarterly Journal of Economics 131*(2), 891–941.

Baicker, K., A. Finkelstein, J. Song, and S. Taubman (2014). The impact of medicaid on labor market activity and program participation: evidence from the oregon health insurance experiment. *American Economic Review 104*(5), 322–28.

Bailey, M. J. and A. Goodman-Bacon (2015). The war on poverty's experiment in public medicine: Community health centers and the mortality of older americans. *American Economic Review 105*(3), 1067–1104.

Beatty, M. E., P. Beutels, M. I. Meltzer, D. S. Shepard, J. Hombach, R. Hutubessy, D. Dessis, L. Coudeville, B. Dervaux, O. Wichmann, et al. (2011). Health economics of dengue: a systematic literature review and expert panel's assessment. *The American journal of tropical medicine and hygiene 84*(3), 473–488.

Bhalotra, S. R., R. Rocha, and R. R. Soares (2019). Does universalization of health work? evidence from health systems restructuring and expansion in brazil. Technical report, IZA Discussion Paper.

Bhatt, S., P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, O. Sankoh, et al. (2013). The global distribution and burden of dengue. *Nature 496*(7446), 504.

Bloom, D. E., D. Canning, et al. (2006). *Epidemics and economics.* na.

Brady, O. J., M. A. Johansson, C. A. Guerra, S. Bhatt, N. Golding, D. M. Pigott, H. Delatte, M. G. Grech, P. T. Leisnham, R. Maciel-de Freitas, et al. (2013). Modelling adult aedes aegypti and aedes albopictus survival at different temperatures in laboratory and field settings. *Parasites & vectors 6*(1), 351.

Currie, J. and B. C. Madrian (1999). Health, health insurance and the labor market. *Handbook of labor economics 3*, 3309–3416.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy.* The World Bank.

Deaton, A. (2003). Health, income, and inequality. *National Bureau of Economic Research Reporter: Research Summary. Retrieved August 15*, 2009.

Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature 52*(3), 740–98.

Deschênes, O. and M. Greenstone (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the us. *American Economic Journal: Applied Economics 3*(4), 152–85.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, K. Baicker, and O. H. S. Group (2012). The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics 127*(3), 1057–1106.

Gertler, P. and J. Gruber (2002). Insuring consumption against illness. *American economic review 92*(1), 51–70.

Goodman-Bacon, A. (2018). Public insurance and mortality: evidence from medicaid implementation. *Journal of Political Economy 126*(1), 216–262.

Graff Zivin, J. and M. Neidell (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics 32*(1), 1–26.

Heal, G. and J. Park (2015). Goldilocks economies? temperature stress and the direct impacts of climate change. Technical report, National Bureau of Economic Research.

Kudamatsu, M., T. Persson, and D. Strömberg (2012). Weather and infant mortality in africa.

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data*, Volume 53. John Wiley & Sons Incorporated.

Limwattananon, S., S. Neelsen, O. O'Donnell, P. Prakongsai, V. Tangcharoensathien, E. Van Doorslaer, and V. Vongmongkol (2015). Universal coverage with supply-side reform: The impact on medical expenditure risk and utilization in thailand. *Journal of Public Economics 121*, 79–94.

Liu-Helmersson, J., H. Stenlund, A. Wilder-Smith, and J. Rocklöv (2014). Vectorial capacity of aedes aegypti: effects of temperature and implications for global dengue epidemic potential. *PloS one 9*(3), e89783.

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

Marinescu, I. (2014). Hiv, wages, and the skill premium. *Journal of health economics 37*, 181–197.

Mazumder, B. and S. Miller (2016). The effects of the massachusetts health reform on household financial distress. *American Economic Journal: Economic Policy 8*(3), 284–313.

Messina, J. P., O. J. Brady, T. W. Scott, C. Zou, D. M. Pigott, K. A. Duda, S. Bhatt, L. Katzelnick, R. E. Howes, K. E. Battle, et al. (2014). Global spread of dengue virus types: mapping the 70 year history. *Trends in microbiology 22*(3), 138–146.

Miller, G., D. Pinto, and M. Vera-Hernández (2013). Risk protection, service use, and health outcomes under colombia's health insurance program for the poor. *American economic journal. Applied economics 5*(4), 61–91.

Obolski, U., P. N. Perez, C. J. Villabona-Arenas, N. R. Faria, and J. Lourenco (2018). Mvse: an r-package that estimates a climate-driven mosquito-borne viral suitability index. *bioRxiv*, 360701.

Panamerican Health Organization (2014). *Report on the Situation of Malaria in the Americas 2014*. Panamerican Health Organization.

Prinz, D., M. Chernew, D. Cutler, and A. Frakt (2018). Health and economic activity over the lifecycle: Literature review. Technical report, National Bureau of Economic Research.

Rocha, R. and R. R. Soares (2010). Evaluating the impact of community-based health interventions: evidence from brazil's family health program. *Health economics 19*(S1), 126–158.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics 6*(2), 461–464.

Shepard, D. S., E. A. Undurraga, Y. A. Halasa, and J. D. Stanaway (2016). The global economic burden of dengue: a systematic analysis. *The Lancet infectious diseases 16*(8), 935–941.

Silva, M. M., M. S. Rodrigues, I. A. Paploski, M. Kikuti, A. M. Kasper, J. S. Cruz, T. L. Queiroz, A. S. Tavares, P. M. Santana, J. M. Araújo, et al. (2016). Accuracy of dengue reporting by national surveillance system, brazil. *Emerging infectious diseases 22*(2), 336.

Suaya, J. A., D. S. Shepard, J. B. Siqueira, C. T. Martelli, L. C. Lum, L. H. Tan, S. Kongsin, S. Jiamton, F. Garrido, R. Montoya, et al. (2009). Cost of dengue cases in eight countries in the americas and asia: a prospective study. *The American journal of tropical medicine and hygiene 80*(5), 846–855.

Teixeira, M. G., J. B. Siqueira Jr, G. L. Ferreira, L. Bricks, and G. Joint (2013). Epidemiological trends of dengue disease in brazil (2000–2010): a systematic literature search and analysis. *PLoS neglected tropical diseases 7*(12), e2520.

Undurraga, E. A., M. Betancourt-Cravioto, J. Ramos-Castañeda, R. Martínez-Vega, J. Méndez-Galván, D. J. Gubler, M. G. Guzmán, S. B. Halstead, E. Harris, P. Kuri-Morales, et al. (2015). Economic and disease burden of dengue in mexico. *PLoS neglected tropical diseases 9*(3), e0003547.

Wagstaff, A. (2007). The economic consequences of health shocks: evidence from vietnam. *Journal of health economics 26*(1), 82–100.

Weil, D. N. (2014). Health and economic growth. In *Handbook of economic growth*, Volume 2, pp. 623–682. Elsevier.

World Health Organization (2009). *Dengue: guidelines for diagnosis, treatment, prevention and control*. World Health Organization.

Xavier, A. C., C. W. King, and B. R. Scanlon (2016). Daily gridded meteorological variables in brazil (1980–2013). *International Journal of Climatology 36*(6), 2644–2659.

Zhang, P., O. Deschenes, K. Meng, and J. Zhang (2018). Temperature effects on productivity and factor reallocation: Evidence from a half million chinese manufacturing plants. *Journal of Environmental Economics and Management 88*, 1–17.

A Appendix Data Sources

A.1 Socioeconomic Indicators and Access to Welfare: - CadUnico

CadUnico is a national registry system, organized by the Ministry of Social Development and Fight Against Hunger (MDS) in Brazil, that aims to map over time all families and individuals of low socioeconomic status in the country. Every family with either a monthly per capita income of at most half the minimum salary or a monthly total income of up to 3 minimum salaries should be registered.

CadUnico is the entry point to reach most of government programs. Families are included in the registry under three different situations: (i) when they sign up for a social assistance program, (ii) when they voluntarily register themselves or (iii) when local governments actively try to register poor families independently of current eligibility status for a specific program. Implemented in 2001, CadUnico has since evolved following the introduction of social programs that incentivize families to register and keep their records up to date. Today, it is the most comprehensive longitudinal socioeconomic profile of over 16 million families and 60 million individuals in the country. However, data available for research do not go back this far, as early records contain quality issues related to the lack of data entry checks and national level consistencies.

The CadUnico registry system is managed through two levels of government. At the local level, municipalities have the responsibility of identifying families in the target group, collecting their information and keeping their records up to date. From time to time, they are required to formally visit at least 10 percent of the families for data verification. At the national level, the federal government through CAIXA, a state-owned bank, manages payroll accounting duties, provides IT services that unify the data entry system, consolidates the database and monitors data quality. The MDS in turn designs CadUnico regulations, oversees and coordinates municipalities and CAIXA's actions.

In practice, the federal government uses information from CadUnico for planning and actively identifying families that are eligible for social assistance programs. About eight federal government social security policies, including those of housing and subsidized landlines, use the registry to identify specific vulnerable groups they aim to reach. The Bolsa Familia Program (BFP) is its main user. BFP is one of the largest CCT program in the world, and has the largest coverage among all social security programs in the country.

Upon registration the family is asked to provide a permanent address and at least one valid identification document for each member of the family. Families are uniquely identified through an 11-digit sequence number, the Number of Social Identification (*Numero de Identificacao Social* – NIS), with which they can be followed over time and be easily identified in other governmental databases.

CadUnico requires families that are part of BFP to update their records each 24 months. If there is any change of circumstances within this period they should voluntarily anticipate it. If the family fails to provide an update when recalled they receive an update recall message in their payroll slip. After 24 months since the last update the municipality should actively try and locate the family to keep track of its current status. The benefit will effectively be canceled approximately 9 months after a family record has been identified as outdated.

Both when they first register and when update their records, the family representative, typically the mother, responds to surveys that profile labor market participation, income, education, disabilities, and enrollment in other social programs for all family members. She also updates the household address, the school identification number at which children are enrolled, and the health facility the family usually access when in need. We rely on these pieces of information to define socioeconomic variables at the household level, to geocode and group family location, and to define indicators of access to social programs.

A.2 Dengue Data - SINAN

Dengue fever is a mandatorily reportable disease in Brazil. All public and private health facilities that have their services contracted by the government are required to do so. Hence, the data cover all municipalities in the country.

Suspected cases are routinely reported by using standardized forms completed by clinicians or health staff and subsequently sent to local health surveillance officials. Data from these forms are entered into the Notifiable Diseases Information System (*Sistema de Informacao de Agravos de Notificacao* – SINAN), a database that compiles all the information on compulsorily notifiable diseases in the country. It is used for monitoring diseases with the potential to become epidemics, to identify their origin and put in place preventive measures. The database includes information on basic demographics, dates of notification and symptom onset, case classification that adheres to PAHO/WHO guidelines, details on symptoms, pre-existing morbidities, and outcomes. The first few cases registered in a given region should receive laboratorial confirmation, after which only clinical diagnosis becomes necessary, except for cases with warning signs for severity.

Failure to transmit data from the local level for a period of two months is penalized by cancellation of financial resources to the municipality. In the case of non-occurrence of a disease, health facilities are still required to fill forms stating just that, so as to avoid underreporting. However, it is believed that reporting is still incomplete. This is due to a number of reasons, ranging from failures in recognizing the disease to technical or administrative issues related to the notification system. Silva et al. (2016) estimate that only 1 in every 12 cases are reported during a typical dengue season, and this number goes up to 17 in periods of low transmission.

B Appendix Subsection 3.2: Macro-effects

B.1 Other Tables and Graphs

Figure B.1: Time series of dengue incidence per 1,000 people by Metropolitan-Region

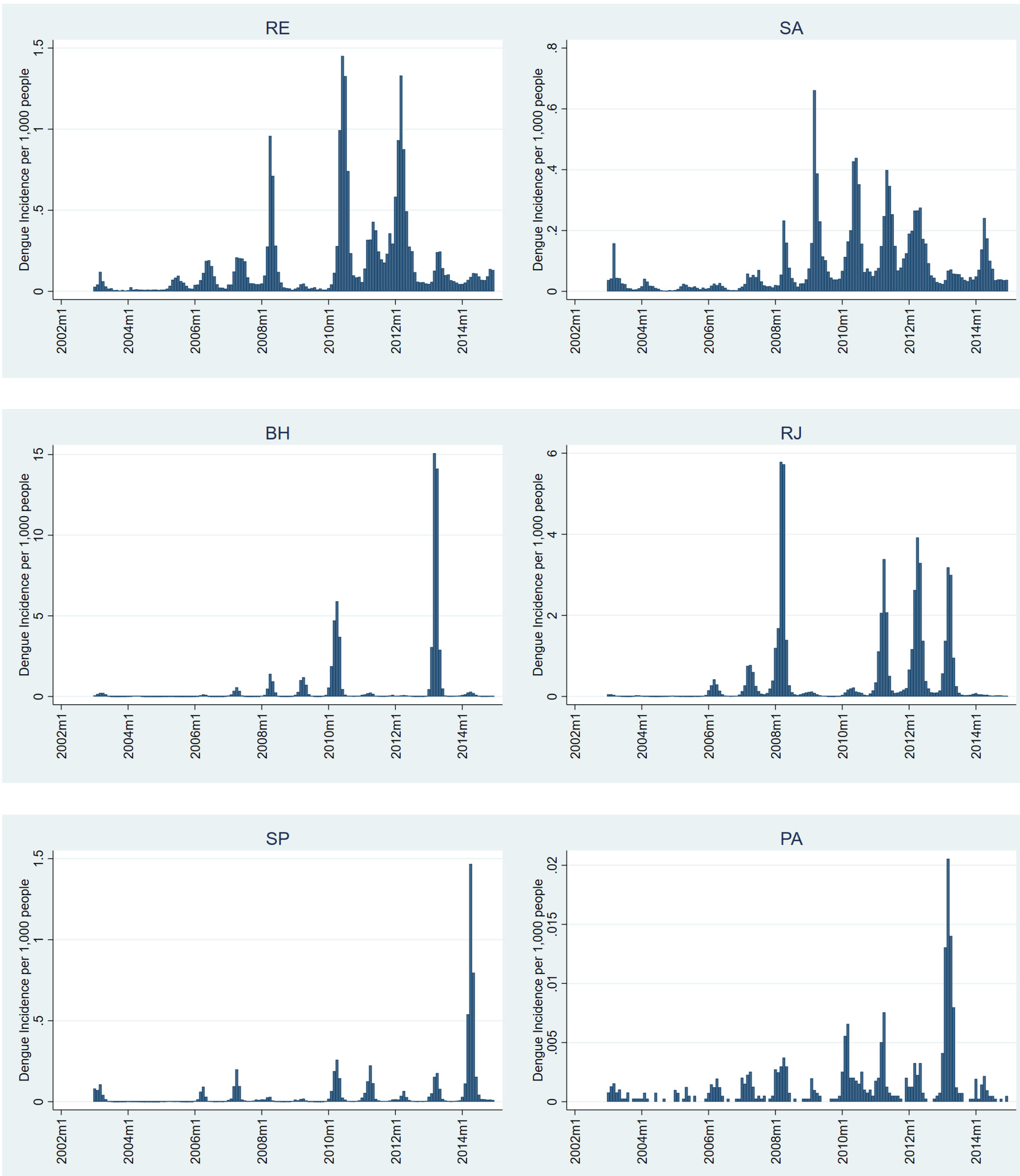


Figure B.2: Time series of daily mean temperature by Metropolitan-Region

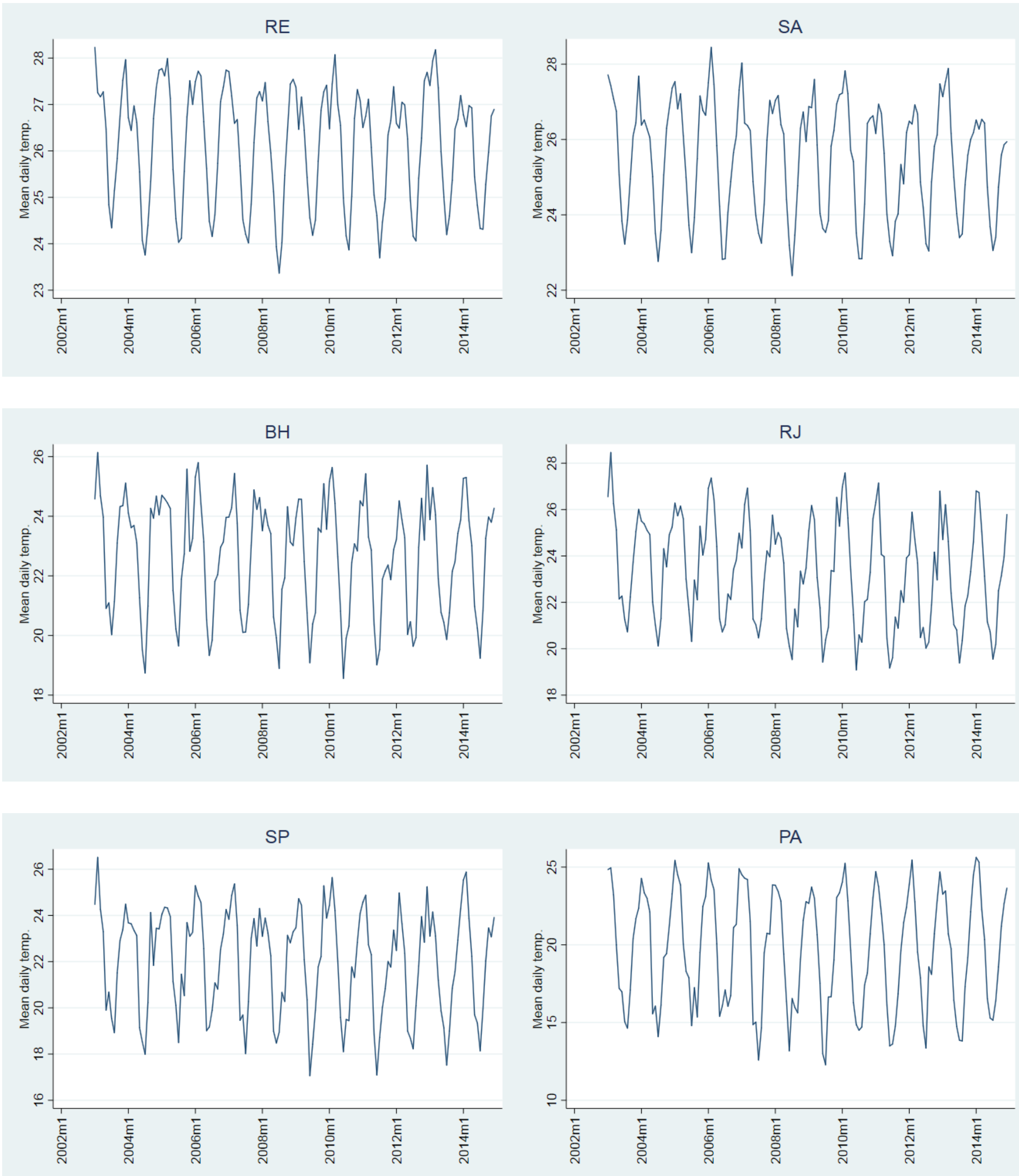


Figure B.3: Time series of daily mean relative-humidity by Metropolitan-Region

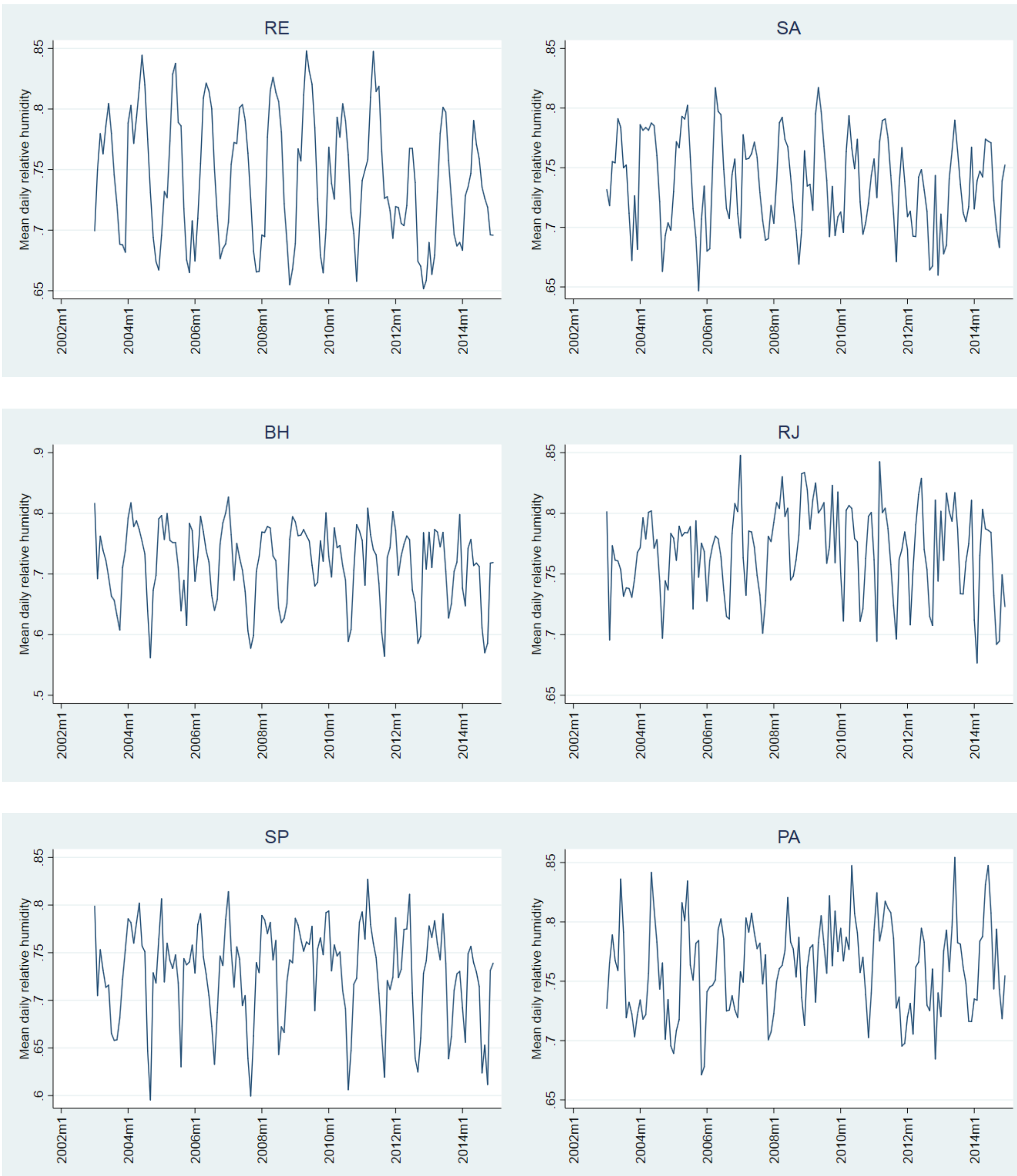


Table B.1: Reduced-form estimates by worker type

	Formal			Informal			Self-employed		
	Leave	Hours	Earn.	Leave	Hours	Earn.	Leave	Hours	Earn.
DSI	0.004***	-0.001	0.001	0.002***	-0.003*	0.000	0.003***	-0.003	-0.007*
	(0.001)	(0.001)	(0.004)	(0.001)	(0.002)	(0.005)	(0.001)	(0.002)	(0.004)
DSI x Post	0.001**	-0.003***	-0.001	0.002***	-0.002	-0.003	0.002***	-0.005***	-0.003
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.005)	(0.001)	(0.001)	(0.004)
Observations	864	864	858	864	864	858	864	864	858
Region FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Month FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Region t-trend	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean dep.	0.0244	3.755	7.487	0.00904	3.672	7.098	0.0120	3.690	7.310

Note: Standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1

Table B.2: Reduced-form estimates by gender

	Male			Female		
	Leave	Hours	Earn.	Leave	Hours	Earn.
DSI	0.002***	-0.001	0.003	0.003***	-0.002	0.001
	(0.000)	(0.001)	(0.003)	(0.000)	(0.001)	(0.004)
DSI x Post	0.001***	-0.003***	-0.005**	0.001**	-0.003***	-0.007***
	(0.000)	(0.001)	(0.002)	(0.000)	(0.001)	(0.003)
Observations	864	864	858	864	864	858
Region FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Month FE	YES	YES	YES	YES	YES	YES
Region t-trend	YES	YES	YES	YES	YES	YES
Mean dep.	0.0149	3.757	7.578	0.0142	3.661	7.411

Note: Standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1

B.2 Summary Statistics

Table B.3: Summary Statistics - Labor Market Outcomes

	2003-2007		2008-2014	
	mean	sd	mean	sd
In labor-force	0.63	0.04	0.64	0.04
Working	0.87	0.04	0.92	0.03
Hours	41.58	0.93	41.04	0.96
On-leave	0.01	0.01	0.01	0.01
Earnings	1,655.78	307.85	2,027.59	354.44
Dengue (GP)	325.96	964.96	2,392.37	8,636.35
DSI	1.18	0.42	1.11	0.40
Observations	360	360	504	504
Metro-Regions	6	6	6	6

Table B.4: Summary Statistics - Morbidity Outcomes

	2003-2007		2008-2014	
	mean	sd	mean	sd
Dengue	0.61	4.61	1.07	9.89
Infectious	11.94	35.72	12.19	41.75
Chronic	76.92	351.33	76.66	369.11
External	13.71	80.88	17.26	99.07
NCD: neoplasms	6.42	40.66	7.44	47.99
NCD: diabetes	1.86	8.40	2.29	8.60
NCD: cardio	18.47	89.78	18.62	96.60
NCD: respiratory	11.09	40.36	8.52	34.56
Dengue (GP)	3.91	65.48	12.06	228.36
DSI	1.20	0.47	1.11	0.43
Observations	321,336	321,336	450,348	450,348
Municipalities	5,359	5,359	5,363	5,363

Table B.5: Summary Statistics - Welfare Outcomes

	2011-2016	
	mean	sd
PBF: Families	2,476	8,055
PBF: Value	380,663	1,141,872
INSS: Dengue-sickleaves	.013	.22
Dengue (GP)	15	264
DSI	1.1	.4
Observations	446,880	446,880
Municipalities	5,320	5,320

B.3 Dengue Suitability Index

The Dengue Suitability Index was developed by (Obolski et al., 2018), based on a vector dynamic-population model developed in previous studies. In this model, there are 4 weather-dependent parameters: (i) life-span of adult mosquitoes (μ); (ii) extrinsic incubation (γ); (iii) daily biting rate (a); and, (iv) probability of transmission from infected mosquito to human per bite (ϕ). Figure B.4 show how each parameter depends on temperature and humidity respectively.

Figure B.4: Weather dependent parameters

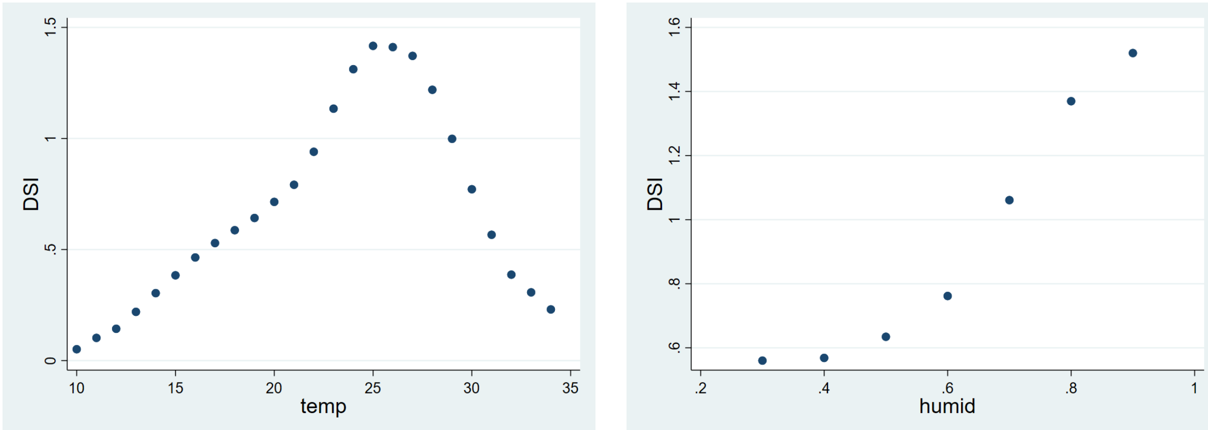
$$\begin{aligned} \breve{\mu}^{v}_{(t)} &= 0.8692 - 0.1599t + 0.01116t^2 - 0.0003408t^3 + 0.000003809t^4 \\ \breve{\gamma}^{v}_{(t)} &= \frac{0.003359\frac{tk}{298} \times \exp(\frac{15000}{R}(\frac{1}{298} - \frac{1}{tk}))}{1 + \exp(\frac{6.203\times10^{21}}{R}(\frac{1}{-2.176\times10^{30}} - \frac{1}{tk}))} \\ \breve{\phi}^{v\to h}_{(t)} &= 0.001044t \times (t - 12.286) \times (32.461 - t)^{1/2} \end{aligned}$$

$$\begin{aligned} \breve{a}^{v}_{(u)} &= (u - \bar{u})/\sqrt{1 + (u - \bar{u})^2} \\ \breve{\mu}^{v}_{(u)} &= \bar{u} - (u - \bar{u})/\sqrt{1 + (u - \bar{u})^2} \end{aligned}$$

Note: Figures taken from Appendix 2.1 of (Obolski et al., 2018)

Figures B.5a and B.5b show the relationship between DSI and temperature and humidity, respectively. We can observe that the relationship seems to be quadratic on temperature and possibly cubic in humidity.

Figure B.5: Relationship between DSI and weather



(a) Temperature (b) Humidity

B.4 Measurement Error

We show in this section how measurement error affects the estimation of equation Equation (1). Denote the observed prevalence of dengue as: $\dot{D}_{it} = D_{it} + \eta_{it}$, where η_{it} is the error arising from the fact that prevalence is estimated from the number of cases detected and reported by a general practitioner. This error will most likely not be classical. In fact, the epidemiological literature normally propose that there is a fix factor of cases that the system will not observe. Below we will discuss different potential data generating processes for this error, and the consequences for our estimates.

B.4.1 ME type 1

$$\dot{D}_{it} = D_{it} + \eta_{it} \tag{5}$$

$$\eta_{it} = \alpha D_{it} \tag{6}$$

where α is a negative scalar between 0 and 1. Replacing (6) in (5) we get $\dot{D}_{it} = (1+\alpha)D_{it}$. This is the most common approach, assuming that a fraction α of the cases are unobserved, and this fraction is constant in time and space.

It is easy to show that in this case the OLS estimate of β in Equation (1) is biased:

$$\hat{\beta}_{OLS} = \frac{cov(y,\dot{D})}{var(\dot{D})} = \frac{cov(y,(1+\alpha)D)}{var((1+\alpha)D)} = \frac{1}{(1+\alpha)}\beta \tag{7}$$

However, if we standardized the prevalence D, the new variable $\tilde{D}$ has mean zero and standard deviation 1, and no measurement error.

$$\tilde{D}_{it} = \frac{\dot{D} - \mu_{\dot{D}}}{\sigma_{\dot{D}}} = \frac{(1+\alpha)D - (1+\alpha)\mu_D}{(1+\alpha)\sigma_D} = \frac{D - \mu_D}{\sigma_D} \sim N(0,1) \tag{8}$$

Under this scenario, using the OLS estimator on $\tilde{D}$ gives an unbiased estimator of β.

A simple extension of this case is when the error term has both a part that is proportional to D and an error term with mean zero and uncorrelated with ϵ and D -classical measurement error. In this case 2SLS estimates are consistent.

B.4.2 ME type 2

$$\dot{D}_{it} = D_{it} + \eta_{it} \tag{9}$$

$$\eta_{it} = \alpha Z_{it} \tag{10}$$

$$cov(Z, \epsilon) \neq 0 \tag{11}$$

where α is a negative scalar between 0 and 1, and Z is an unobserved variable correlated with the error term in Equation (1). Replacing (10) in (9) we get $\dot{D}_{it} = D_{it} + \alpha Z_{it}$. In this scenario the under-reporting is correlated with some characteristic of the region (e.g. the share of poor households, if poor individuals are less likely to see a physician when feeling sick).

Because Z is correlated with both $\dot{D}$ and ϵ, OLS estimates would suffer from the usual omitted variable bias. However, using a 2SLS estimator with the DSI as instrument gives a consistent estimate of β.

C Appendix Subsection 3.3: Micro-effects

C.1 Summary Statistics

Table C.1

	Sample: 2016 (baseline), only head of HHs				
	(1)	(2)	(3)	(4)	(5)
	Municipalities with at least 1 case of dengue in 2017				
	HHs Without Dengue		HHs With Dengue		Diff
	mean	sd	mean	sd	
% HH connected to water supply	0.803	0.397	0.845	0.362	0.042***
% HH connected to sanitation network	0.397	0.489	0.354	0.478	-0.043***
% HH connected to electricity newtwork	0.801	0.399	0.860	0.347	0.059***
Average Highest School Grade Completed	2.483	2.606	3.116	2.625	0.633***
% Non-White	0.802	0.398	0.816	0.388	0.014**
% Female	0.578	0.494	0.632	0.482	0.054***
	Zipcodes with at least 1 case of dengue in 2017				
	HHs Without Dengue		HHs With Dengue		Diff
	mean	sd	mean	sd	
% HH connected to water supply	0.745	0.436	0.845	0.362	0.100***
% HH connected to sanitation network	0.265	0.441	0.354	0.478	0.089***
% HH connected to electricity newtwork	0.822	0.383	0.860	0.347	0.038***
Average Highest School Grade Completed	2.462	2.593	3.116	2.625	0.654***
% Non-White	0.827	0.378	0.816	0.388	-0.011*
% Female	0.565	0.496	0.632	0.482	0.067***
	Schools with at least 1 case of dengue in 2017				
	HHs Without Dengue		HHs With Dengue		Diff
	mean	sd	mean	sd	
% HH connected to water supply	0.828	0.377	0.845	0.362	0.017***
% HH connected to sanitation network	0.429	0.495	0.354	0.478	-0.075***
% HH connected to electricity newtwork	0.814	0.389	0.860	0.347	0.046***
Average Highest School Grade Completed	2.617	2.577	3.116	2.625	0.499***
% Non-White	0.794	0.404	0.816	0.388	0.022***
% Female	0.590	0.492	0.632	0.482	0.041***

Notes: Socioeconomic indicators from CadUnico 2016, indicator of dengue infection from SINAN 2017. In the upper panel, the sample is restricted to households located in municipalities with at least one reported case of dengue in 2017; in the mid-panel, restricted to households located in zipcodes with at least one reported case of dengue; in the bottom panel, restricted to households with children enrolled in schools for which we identify children or any of their respective family members that contracted dengue in the 2017 season.Significance: ***p<0.01, **p<0.05, *p<0.1.