From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

Doctoral Dissertations, 2011-2026 | Public and International Affairs

2018

Quantifying the socioeconomic implications of climate change: Three essays

Baker, Rachel

Princeton, NJ : Princeton University

http://arks.princeton.edu/ark:/88435/dsp012514np181

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Quantifying the socioeconomic implications of climate change: Three essays

Rachel Elizabeth Baker

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the
The Woodrow Wilson School of Public and International Affairs
Adviser: Michael Oppenheimer

April 2018

© Copyright by Rachel Elizabeth Baker, 2018.

All rights reserved.

Abstract

Efforts to quantify the future socioeconomic effects of climate change are important for present-day decision-making. Generating these predictions requires nuanced modeling approaches that accurately capture the complexity of the socio-environmental system whilst being grounded in real-world data. This thesis evaluates the effect of climate change on three different outcomes by combining large socioeconomic and climatic datasets, with modeling tools developed by different disciplines.

In the first study I look at how climatic changes may alter HIV prevalence through behavioral responses to poor agricultural years. Statistical relationships are found between warmer periods and behaviors that increase risk of HIV. Using results from approximately 400,000 individuals tested for HIV, I find that the probability of testing positive increases after multiple years of warming. A mechanistic model, parameterized with this statistical data, is used to investigate the implications of this association for long-run changes to the climate.

In the second study I investigate the effect of climate on varicella, a directly-transmitted infection. Laboratory studies for similar diseases have shown a relationship between humidity and transmission. With co-authors, I develop a novel methodology for testing the effect of climatic changes on the transmission rate for varicella that combines statistical tools with a mechanistic model to capture known dynamics of this disease. I find a significant relationship between drier days and increases in varicella transmission. The mechanistic model is used to simulate the effect of climate change on varicella incidence in Mexico.

In the final study I look at how temperature affects labor supply using an employment survey from six urban areas in Brazil. Labor supply changes are hypothesized to be one of the causal mechanisms driving the well-studied linkage between the climate and economic growth. I find an association between hot weeks and decreased work time, particularly in high risk sectors such as agriculture and mining. Using

this estimated association, I project the total lost working time as climate change drives up temperatures in this region. I find that the coupled effect of urban heat islands and climate change may result in large losses in work-hours, unless cooling technologies are implemented.

Acknowledgements

Many wonderful people helped make this thesis possible. First, I would like to thank my adviser, Michael Oppenheimer, for providing guidance throughout the PhD, for encouraging me to explore and develop my own ideas and for providing honest and generous feedback on those ideas, as they developed. It has been a privilege and a joy to learn so much about this field from him. I would also like to thank Tom Vogl, whose brilliant class in econometrics inspired the methodological direction of this thesis. Furthermore, by assigning Dell, Jones and Olken as a reading in his class, he (unknowingly) introduced me to the empirical climate-impacts field, in which all my work has been situated. I'm very glad to have had his feedback as a member of my committee.

A big thanks goes to Solomon Hsiang. First, for allowing me to join the excellent Climate Econometrics summercamp, where I got a real grounding in the specific methods of the climate-impacts field. Second, for allowing me to visit his lab group for a year. Sol has been a great role-model in terms of how to be a quantitatively-driven, interdisciplinary researcher. He has inspired me to not be afraid of asking the big questions, and not to fear methods that differ from my own. I would also like to thank Sol's lab group: Tamma Carleton, Jon Proctor, Jon Kadish, Andy Hultgren, Peiley Lau – their intellectual and motivational support have been invaluable.

Similarly, I would like to thank the STEP PhD community, particularly, David Kanter, Yue Qin, Philip Hannam, Wei Peng, Maya Buchanan, Mark Walker, Mayank Misra, DJ Rasmussen and many others. I have learnt so much from being surrounded by such kind, interesting and interdisciplinary peers. To be able to discuss ideas, whether formally or informally, with this group of people, has been a great joy of my time at Princeton. In addition, Chuck Crosby makes the STEP community what it is: thanks for the tea, cookies and laptops, when I needed them all.

Princeton has some great forums for cross-disciplinary discussion. I would like to thank the CEREAL (Conversations on Environment, Renewable Energy, And Life) breakfast group, particularly Holly Welles and Rob Socolow, for providing such a forum, as well as Princeton Energy and Climate Scholars (PECS).

I would like to thank Ayesha Mahmud and Jessica Metcalf for aiding my introduction to the climate-disease field. In particular, Ayesha has spent many hours answering my questions about disease models with enduring patience. I am grateful to both Jess and Ayesha for their co-authorship on my third chapter and their willingness to discuss further ideas. I hope we have many more papers together to come.

I would like to thank my family for their love and support. In particular, my mum for helping me with questions of grammar; my dad for encouraging my love of mathematics and coding; and my sister, for being a brilliant female academic role-model. Finally, I would like to thank my great friends: CC, Zara and Ayesha, who helped make this place feel like home.

"The sun flared down on the growing corn day after day until a line of brown spread along the edge of each green bayonet. The clouds appeared, and went away, and in a while they did not try any more. The weeds grew darker green to protect themselves, and they did not spread any more. The surface of the earth crusted, a thin hard crust, and as the sky became pale, so the earth became pale, pink in the red country and white in the gray country." – The Grapes of Wrath

Contents

Abstract iii

Acknowledgements v

List of Tables xi

List of Figures xii

1 Introduction **1**

1.1 Why make climate change predictions? 3

1.1.1 Social cost of carbon 4

1.1.2 Adaptation planning 6

1.2 Incorporating Data 7

1.2.1 The impact of weather variation on socio-economic outcomes . 7

1.2.2 Predicting economic outcomes of climate change 10

1.3 Data-driven approaches for predicting climate change impacts 12

1.3.1 Examples of hybrid modeling approaches 15

1.4 Outline of the thesis 18

2 Climate change drives modeled HIV prevalence **21**

2.1 Introduction 22

2.2 Data 24

2.3 Methods 24

2.4 Results 28

2.4.1 Current response of HIV to temperature changes 28
2.4.2 Modeling the future response of HIV prevalence to climate change 32
2.5 Discussion 37

3 Dynamic response of airborne infections to climate change: predictions for varicella 39

3.1 Introduction 40
3.2 Data 43
3.3 Methods 44
3.3.1 The TSIR model 44
3.3.2 Panel regression model 46
3.3.3 Robustness check on simulated data 47
3.3.4 Comparison with other methods 48
3.3.5 Projections 51
3.4 Results 52
3.4.1 Causal Estimation 52
3.4.2 Projections 55
3.5 Discussion 58

4 Implications of climate change for urban labor supply 60

4.1 Introduction 61
4.1.1 Global Climate Prospectus 63
4.2 Data 64
4.3 Methods 67
4.3.1 Temperature driven changes to labor supply 67
4.3.2 Projecting the consequences of climate change 67
4.3.3 Adjusting responses to account for urban heat islands 69
4.4 Results 71

4.4.1 The effect of temperature on labor supply in Brazil 71
4.4.2 Projected changes and costs of changes to labor supply driven by climate change 75
4.4.3 Adjustments for urban heat islands 78
4.5 Discussion 81

5 Conclusion **85**

A Climate change drives modeled HIV prevalence **93**
A.1 Modeling HIV 93
A.1.1 Bayesian MCMC 95
A.2 Tables and Figures 96

B Dynamic response of airborne infections to climate change: predictions for varicella **108**

C Implications of climate change for urban labor supply **121**

Bibliography **132**

List of Tables

2.1 Suggested increase in antiretroviral coverage to offset the potential effects of climate change. 36

A.1 Robustness check: alternative specifications. 97

A.2 Robustness check: tests for misspecification using leads. 101

A.3 Percentage change in HIV prevalence under the RCP8.5 warming scenario. 103

A.4 The output of the model-fitting procedure for each country. 104

B.1 Robustness check: regression results showing the effect of absolute humidity on log transmission. 109

C.1 Model input: efficacy of cooling policies in terms of Celsius reduction. 122

C.2 Model input: projected population growth. 122

C.3 Model input: city-level statistics. 122

C.4 Robustness check: selection tests using the non-parametric binned model. 123

C.5 Robustness check: F-test results (Prob >F) for fourth order polynomial regressions run with different clusterings of standard errors. 126

List of Figures

1.1 Timeline of climate-impacts studies. 3

1.2 Modelling approaches within climate-impacts studies 13

2.1 The effect of temperature changes on HIV prevalence for different ages. 29

2.2 The relationships between temperature and HIV prevalence for different sub-populations. 30

2.3 The effect of temperature changes on risk factors for HIV. 31

2.4 The model results for every country in Sub-Saharan Africa where both HIV and antiretroviral data is available. 34

3.1 Relationship between average relative humidity and the estimated basic reproduction number, R_0. 41

3.2 Estimated effect of humidity and temperature on varicella transmission. 52

3.3 Robustness check of the estimation method on simulated incidence data. 54

3.4 Projections of the impact of climate change on varicella incidence based on 500 simulation runs. 56

3.5 Percentage change in monthly incidence for each state for the years 2070 - 2100 relative to the years 1985-2015. 57

4.1 BEST data and urban heat island differential. 70

4.2 Main effect of temperature supply on working time. 72

4.3 The effect of maximum daily temperature on working time for different employment groups. 73
4.4 Separate fourth-order polynomial response functions fitted to each city by sector. 74
4.5 The projected effect of climate change on change in minutes worked for an individual living in each city. 76
4.6 The projected economic effect of climate change due to changing labor supply across the six study cities. 77
4.7 Cost benefit analysis of city level cooling policies 80
4.8 Cumulative benefits of city cooling policies. 81

A.1 The potential pathways linking temperature changes to HIV prevalence changes. 97
A.2 Data: the locations of Demographic and Health Survey sites. 98
A.3 Data: the number of observations, country and timing of each Demographic and Health Survey. 99
A.4 The effect of temperature changes on HIV prevalence with temperature averaged over different time periods. 100
A.5 Robustness check: polynomial response of HIV prevalence to temperature. 102
A.6 A diagram representing the HIV model. 102
A.7 Modeled incidence between the present day and 2050 over a range of population growth rate scenarios. 105
A.8 The long-run effect of climate change on HIV prevalence under the RCP2.6 scenario. 106
A.9 A visual example of the process for estimating the relationship between temperature changes and p_H for Tanzania. 107

B.1 Data: reported cases of varicella for each state in Mexico from 1985 to 2016. 110
B.2 Data: relative humidity (%) time series for each state in Mexico from 1985 to 2016. 110
B.3 Data: temperature time series (degree C) for each state in Mexico from 1985 to 2016. 111
B.4 Data: the distributions of temperature and relative humidity across all states in Mexico between 1985 to 2015. 112
B.5 Seasonal trend in transmission rates estimated by fitting the TSIR model. 113
B.6 Robustness check: estimate effect of humidity and temperature on varicella transmission assuming an exponential relationship. 114
B.7 Robustness check: estimated effect of humidity on varicella transmission assuming a log-linear relationship between humidity and transmission. 115
B.8 Robustness check: estimated effect of humidity on varicella transmission assuming an exponential relationship between humidity and transmission. 116
B.9 Examples of simulated varicella incidence for four states in Mexico. . 117
B.10 Robustness check: applying the model to simulated incidence data . . 118
B.11 Robustness check: testing alternate specifications of the panel model. 119
B.12 Robustness check: the effect of humidity at different time steps on the transmission rate. 120

C.1 Data: the locations of the 6 cities used in the Pesquisa Mensal de Emprego employment survey. 121
C.2 Data: trends in working hours over time for each city. 124
C.3 Data: seasonal patterns of work for each city. 124
C.4 Data: CMIP5 temperature projections for each city. 125

C.5 Data: the ratio of primary, secondary and tertiary sector employment. 125
C.6 The cumulative benefits accrued from implementing each of three cooling policy options. 126
C.7 Robustness check: plot of p-values from selection tests. 127
C.8 Robustness check: tests for misspecification using leads. 128
C.9 Robustness check: tests for misspecification using lags. 129
C.10 The UHI-adjusted response of labor supply to temperature changes. . 130
C.11 The link between economic income and temperature changes to working hours. 131

Chapter 1

Introduction

Climate change presents one of the greatest challenges our species has known. With diminishing hope in a robust international environmental agreement to reduce greenhouse gas emissions, and the repercussions of extreme weather events already having a significant impact on society, attention is turning from mitigation to potential adaptation options given that climatic changes are inevitable.[1]. Prediction, both in terms of the physical realizations of future changes, as well as their associated effect on human society, is a crucial part of this adaptation planning.

Predicting the future effect of climate change on society presents a novel challenge for researchers and policy makers. Unlike other prediction problems, some of the outcomes of climate change may occur decades, if not centuries, in the future. Predictions also may depict a reality that with appropriate action will not come to pass. In this sense, predictions are endogenous in that they may lead to actions, such as adaptation, that diminish the probability of predicted outcomes occurring.

This long-term nature, and the potential endogeneity, makes socioeconomic climate predictions a unique problem from a modeling perspective. In other disciplines

[1]Present-day extreme weather events are attributed to climate change (or otherwise) using probabilistic methods. It is impossible to say for certain whether any particular event was *caused* by climate change, but the *fraction of attributable risk* gives the fraction of the current risk that is attributable to greenhouse gas emission [129]

predictions are validated by repeated experiments, by withholding some portion of the data for testing, or by observing near-term realizations. In contrast, the first wave of Integrated Assessment Models (IAMs), used to estimate the economic effects of climate change, contained functions "with no theoretical or empirical foundation" [111]. The lack of empirical basis has lead to wide spread criticism and uncertainty in these modeled results [89].

The "new climate-economy literature" has been developed in recent years to incorporate data-driven analysis into climate-impacts research [27]. This literature uses econometric methods to find associations between (generally short-run) climatic variations and economic or other social outcomes. Climate datasets are combined with socioeconomic datasets, usually with some degree of spatial and temporal variation. Econometric methods allow the identification of plausibly causal associations between weather shocks and outcomes.

Claims of causality can be controversial within scientific research [107]. The elegance of econometric methods lies in appropriately testing causal relationships. In the climate-impacts case, the exogenous nature of (most) weather variation facilitates causal claims. In this Introduction I shall review briefly the methodology of such studies–these methods being the foundation of the following three chapters in this thesis. I shall also consider how such studies can be used as grounds for predicting the future impact of climate change on the economy. Finally, I shall consider the limitations of such work, particularly for some sectors of outcomes, and how this thesis attempts to overcome them.

Data-driven climate-impacts work typically uses datasets covering the recent past to generate predictions over the short-term and long-term future. The uncertainty in the prediction increases with time. Typically, climate predictions hinge around the midpoint of the century, 2050, and the end of the century, 2100. Figure 1.1 shows these key time points, relative to the present. The recent past covers time

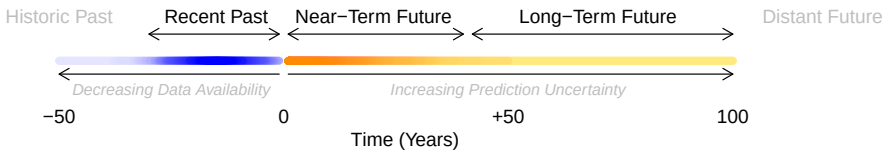


Figure 1.1: A simple reference timeline for climate impacts studies. Data on socioeconomic outcomes typically ranges between 0 and 30 years in the past. Predictions are usually generated to 2050 or 2100. Data availability decreases in negative time, and prediction ability decreases in positive time.

from approximately 30 years ago to the present and generally refers to the era over which the advent of the internet has made the supply, collection and storage of high resolution socio-economic datasets increasingly possible. Most climate-impact studies use socioeconomic datasets that span some portion of this period, though datasets on metrics such as economic growth that are typically collected at the annual level, cover longer stretches of time. The near-term future is considered as the next 30-40 years, up until 2050. Predictions over this length of time are based on a world that is relatively similar to the present day. The long-term future is considered as the following 100 years. Generally predictions do not go beyond this point [129], but some have urged for increasing consideration of events even within the distant future [127].

1.1 Why make climate change predictions?

There are many reasons why policy makers may need predictions for the future effect of climate change. Within the US context, estimates of the *social cost of carbon* rely on assumptions regarding future economic damages from climate change and are core

to environmental policy making. In other realms of policy-making, predictions may allow national or sub-national governments to shape adaptation plans.

1.1.1 Social cost of carbon

The use of cost-benefit analysis in U.S. environmental policy making has its roots in methods originally developed by the RAND cooperation to deal with risks in nuclear weapons engineering. A migration of employees between the RAND cooperation, major universities and the government, around the 1970s, brought these quantitative analysis techniques to the forefront of US policy making at a time when environmental issues were of increasing concern. The appeal of cost-benefit analyses is that they presented (at least in theory) a value-neutral objective approach to weighing up the pros and cons of enacting a specific policy [93].

In terms of the climate change problem, the costs of reducing greenhouse gases such as potential changes to economic output, or investing in cleaner technology, are weighed against the potential benefit in terms of mitigating the negative consequences of climate change. The social cost of carbon was developed as a measure of the latter part of this calculus: a monetized quantification of incremental climate change damages or the "the change in the discounted value of economic welfare from an additional unit of CO2-equivalent emissions" [98]. To date, the social cost of carbon has been used in almost fifty US federal regulations, including rules and proposals for appliances, transportation, industry, and power generation [32].

The US government developed an interagency working group to assess the social cost of carbon. This group took as input the results of three integrated assessment models (IAMs) [53]. IAMs compute the social cost of carbon by inputting greenhouse gas emission pathways and estimating the total economic effect (global or local) of a 1 tonne change in carbon dioxide (equivalent) emitted. The approach rests on several parameters and modeling choices. Three key pieces are the *climate sensitivity*, which

converts changes in greenhouse gas concentrations to changes in temperatures[2]; the *damage function*, which converts changes in temperature to monetized damages; and the *discount rate*, which applies a weight to future economic losses such that economic consequences in future times are valued proportionally less than the present.

These models have been criticized on multiple fronts [69, 111, 89, 118]. The lack of empirical support for the damage function has, in particular, been a focus of potential model improvement [111]. Of the three IAMs used by the EPA, DICE uses a simple quadratic function of temperature to specify damages whereas FUND uses separate damages based on responses of 14 different sectors. PAGE evaluates economic and non-economic damages separately and applies a separate function for damages arising from sea-level rise [32].

Climate-impact studies can be used to update these damage functions by improving the functional form of relationships between temperature changes and the outcomes of interest. The Global Climate Prospectus, to which part of the work within this thesis will contribute, was set up with the specific aim of improving the damage function within IAMs [7]. The approach involves assessing economic damages across multiple sectors and synthesizing these damages within integrated assessment models. In this thesis, contributions are made to one aspect of this calculation: the effect of temperature on global labor supply. However, to truly account for the global cost of carbon, multiple other types of damages and sectors should be taken into account such as changes to morbidity, mortality, migration, industry and agriculture. Quantifying these effects across all regions and all sectors requires a large modeling effort and necessarily involves large uncertainties; yet the alternative approach of removing data from the process makes model results essentially guesswork.

It should be noted that alternatives exist to a bottom-up analysis of the social cost of carbon. In 2009 the UK switched from using a social cost of carbon to a "target-

[2]FUND [4], DICE [97] , and PAGE [61], the three models used by the EPA, all use the climate sensitivity to estimate temperature changes from emissions.

consistent" cost of carbon. This approach suggests that the price of carbon should be based on the marginal abatement curve such that policies are optimized to meet an external greenhouse gas target. Part of the justification for the UK government's switch to a target-based approach is the uncertain nature of the damage function within IAMs. However, the target-based approach is also based on expert elicitation that concluded that "in order to minimise the risks of dangerous climate change, the central estimate (i.e. 50/50 probability) of temperature rise in 2100 should be kept as close as possible to 2C above pre-industrial levels and that the probability of 4C should be kept to very low levels (below 1%)" [28, 22]. The preference for optimizing the avoidance of dangerous climate change has a long history within international environmental policy making, being a key components of the UN's Framework Convention on Climate Change [115, 101, 102].

1.1.2 Adaptation planning

Efforts to calculate the social cost of carbon, irrespective of the methodology, are used to justify emissions reduction policies in order to mitigate climate change and it's impacts. Yet some of these impacts are already being experienced by society [109, 108, 59, 58], and will continue to be experienced over the coming decades. Efforts to quantify potential future events allow policy makers to prepare for these outcomes and adapt accordingly [41].

Without predictions of the impacts of climate change, adaptations to climatic extremes may only occur *ex post*. For example, measures to reduce the effect of heat waves on mortality were introduced in European cities only after the 2003 heat wave that killed approximately 15,000 [40, 113]. Similarly, improvements to coastal barriers that protect cities from storm surges often occur after a flooding event. Improvements to coastal protection in New Orleans were only made after Hurricane Katrina in 2005– one of the costliest natural disasters in US history [75].

Appropriate predictions of climate change impacts may highlight the need for *ex ante* investment in adaptation at sub-national or otherwise levels. Cost-benefit analyses may still drive such policy decisions, but the scale of the calculation, may be simpler than the efforts required for a global social cost of carbon.

1.2 Incorporating Data

As mentioned above, early integrated assessment models, used to calculate the potential economic outcomes of climate change, were not based on empirical findings. The climate-economy literature was developed to improve the empirical grounding of these models. This literature assesses the impact of (recent) historic variations in weather on prior socio-economic outcomes. These findings can then be used to calibrate predictions of future changes.

1.2.1 The impact of weather variation on socio-economic outcomes

Several excellent reviews of the methodology and findings of the climate-impacts field have been published in recent years [21, 27, 62]. As such, I shall only briefly discuss the general approach. Studies typically use both temporal t and spatial variation i to estimate the effect of some climate variable $C_{i,t}$ on some outcome $Y_{i,t}$. Regressions often take the form:

$$Y_{i,t} = \beta f(C_{i,t}) + p(i) + q(t) + \epsilon_{i,t}. \tag{1.1}$$

Climate variables, $C_{i,t}$, include humidity, precipitation, temperature, wind speed and others and can be expresssed with some functional form $f(C_{i,t})$. Outcome variables, $Y_{i,t}$, include crop yields, conflicts, GDP, mortality and others [120, 64, 26, 30]. $p(i)$ and $q(t)$ are controls for space and time respectively, discussed later, and $\epsilon_{i,t}$ is an

the error term. One advantage to this approach from a causal perspective is that climate variables are exogenous — the values that they take are not caused by changes in other variables within the system. For instance, changing yields do not determine the temperature. This property allows for concerns over omitted variable bias and reverse causality, that trouble other applied economics research, to be largely ignored in climate-impact studies. There are three exceptions, though, that must be addressed:

Correlations amongst climate variables. Although we may be able to ignore omitted variables concerns with regard to socioeconomic variables potentially causing changes to climate, there are still known correlations within the climate system itself. Precipitation and temperature have been shown to be correlated and anti-correlated in different locations, for example [6]. Typically, this is resolved by controlling for both variables in the regression. However, as the number of potentially correlated variables grows this approach will increasingly lead to model overfitting. Multi-collinearity is also a concern, as there are multiple similar mappings between aspects of the climate and climate variables.

Time The expected climate is a function of the season and historical time.[3] This in itself does not bias regression results. However, many socio-economic variables may also exhibit seasonal patterns that may be spuriously correlated with the weather. Religious festivals, school holidays, etc. can cause distinct patterns of behavior at certain times of the year which may bias estimated climate effects if not accurately controlled for. Furthermore, long-run changes to climate may be spuriously correlated with long-run changes in economic variables. As such, temporal changes are typically controlled for using some function $q(t)$, often dummy variables for month of year

[3]Clearly this is dependent on the resolution of the climate variable of interest. Annual averages are not affected by seasonality.

and year, but also sinusoidal trends or linear trends. This means that identification comes from random variation in the temporal conditional distribution of climatic variables.

Space Expected climate, as well as socioeconomic outcomes, are often a function of location. Spurious correlation spatially may cause bias in estimates if location effects are not controlled for. $p(i)$ in the Equation 1.1 refers to potential spatial controls which are typically location-specific fixed effects. In the long-run, mean climate in a region may be a driver of local economic effects, although this is very difficult to prove causally.

Tensions over the appropriate time and space controls remains a feature within this literature [16, 15, 65]. However, careful analysis of the data, and robustness checks, can usually elucidate the appropriate model. With corrections for time and space, econometricians can assume *unit homogeneity*: essentially that observations within the data, after conditioning on appropriate controls, are comparable [62]. This assumption is necessary because we never observe the same location, at the same time period, exposed to different weather variation, so we must compare somehow across space and time. Given unit homogeneity, results from regressions using the methodology above can be considered causal.

Studies using the methodology described are able to identify causal associations between climatic changes and socioeconomic outcomes. These results provide some indication of the effect of future climate change on various social or economic sectors. Yet applying these results to predict the long-run impacts of climate change still requires some careful modeling choices.

1.2.2 Predicting economic outcomes of climate change

The output of a climate-impacts study using weather variation is a model of the dependence of Y on C. If Y was monetized (not necessarily an easy task), such a function could be slotted directly within an Integrated Assessment Model to generate economic forecasts. However, applying results from short-run variations in weather to draw conclusions regarding the near-term or long-term change in climate may be problematic for several reasons.

Treatment comparability Human beings may respond very differently to a year-to-year weather shock compared to a long-term change in climate mean, therefore a common critique of this literature is that results from weather changes are not comparable to results from a changing climate. As climate changes permanently, human beings may update their beliefs about the expected climate, and undertake adaptations to minimize negative outcomes. One way to test this effect, suggested by Hsiang 2016 [62], is to estimate models where $f(C_{i,t})$ is evaluated over varying temporal ranges. If responses appear equivalent then the assumption that long-term changes to climate are comparable to short-term changes appears valid. Hsiang defines this as the *marginal treatment comparability assumption* and provides some evidence that in certain circumstances, for certain outcomes, the assumption can be proven to hold.

Threshold effects The marginal treatment comparability assumption, by definition, holds if treatment is marginal, such that small deviations in weather from expected climate can be used to estimate the effect of small deviations in climate. This assumption may hold for the near-term future, but climate changes in the long-term may well be non-marginal. This becomes an issue when climate change produces realizations outside the distribution of historic climate, i.e. climate extremes. These extreme realizations may result in nonlinear responses that are essentially unpre-

dictable based on historic data alone. For instance, a particulary severe heatwave may result in the mass purchase and use of air conditioning units that can cause a surge in energy demand. Sea-level rise coupled with storm surges may cause unexpected mass migration from coastal regions. Predicting the impact of such tail events may be of import to public policy makers due to the potential severity of the consequences, and yet models based on statistical techniques alone are of little use in these cases [131].

Long-term adaptation Changes to local conditions in the long-term, such as economic income and population size, may alter the shape of the climate-impact response function. For instance, as regions increase their economic income they may be able to invest more in adaptation. One novel technique to address these concerns is to estimate response functions conditional on observable covariates. This is achieved relatively easily by interacting $f(C_{i,t})$ with data on economic growth, long-run climatology and population density at a location. Once $f(C_{i,t}, X_i)$ is estimated, where X_i represents the covariates to be interacted with climate, the response function can be combined with future projections of population, economic growth and baseline temperature to predict changes into the long-term future. Essentially this means that response functions from more developed economies that are observed in the present, may be applied to lesser developed economies in the future as they develop. Similar external validity critiques may be levied against this approach as above: i.e. it is difficult to predict outside the range of present-day X and present day C.

Dynamic systems Estimating different response functions based on observable covariates is one option for providing more suitable inputs for prediction. Often the covariates used in X are location-specific and time-invariant within the recent-past

dataset used for estimation. However, time-varying covariates may also shape the response function such that $f(C_{i,t}, X_{i,t})$. Furthermore these time-varying covariates may be dependent on previous states of the system which are determined by previous climate, such that $f(C_{i,t}, X_{i,t}(f(C_{i,t-1}, X_{i,t-1})))$. Such dependencies can be complex and disentangling causal relationships within such a system may require the development of new approaches [31]. In such systems, simple linear models may not uncover the true effect [130].

In summary, predictions of climate change impacts, based on data-driven analysis, are subject to criticism over the extent to which they are robust to changes in the climate system and changes to society, both of which will occur in the future. The extent to which these criticisms can be levied depends very much on the outcome of interest. For example, laboratory experiments may be able to tell us how a particular crop will respond to changing climate, even outside the range that we have historically experienced; however, no experiment can predict the advent of new technology that might improve the robustness of agriculture in general. Robustness along these two dimensions–changing climate and changing society–is crucial in order for a model to generate useful predictions.

1.3 Data-driven approaches for predicting climate change impacts

The difficulty in predictive modeling of climate change impacts lies in balancing data-driven insights with appropriate methods to account for the out-of-sample nature of the problem at hand. I advocate for a hybrid statistical-mechanistic approach to modeling climate change impacts that incorporates empirical analysis with an understanding of the underlying structure of the system involved. The exact approach

of any hybrid statistical-mechanistic model will depend on the nature of the climate-impact interaction and the resolution of the predictions to be generated.

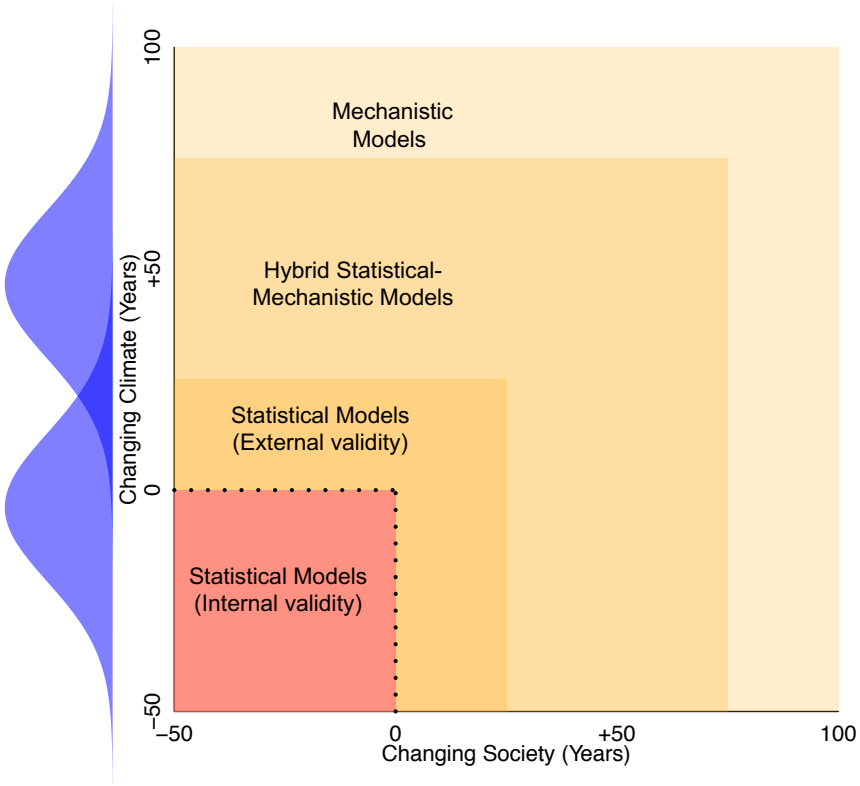


Figure 1.2: A schematic to represent potential optimal modeling strategies for climate-impacts work. The x-axis represents societal changes that a model must be able to incorporate. The y-axis represents changes to the climate. This change is also measured in time representing the historically experienced climate versus the climate that will be experienced in the future. Probabilistic climate distributions are represented stylistically in blue–demonstrating that at more distant time points the climate may change significantly from the present day.

Figure 1.2 gives a schematic of potential modeling approaches based on climate changes and social changes. Moving left to right along the x-axis represents future changes to the society in which climate impacts are occurring: changes to economic income, technology, population distribution, population size are all possibilities. The y-axis considers the climate as a distribution of potential realizations. As time progresses the climate differs increasingly from the current distribution such that at some point events are experienced that have a close to zero probability given the current

distribution. Datasets used in climate-impact studies are represented in terms of climate and social covariates by the red square. Causal relationships assessed using these datasets are valid for the spatial, temporal and climatic range of the dataset.

Near-term future time periods may reflect a society that is fundamentally very similar to that of today. For instance, the effect of hot days on corn yields in 2019 is expected to be very similar to the effect of hot days in 2017. Similarly, the effect of climate distributions that are marginally different from today's distribution may be modeled using responses from the current distribution: most draws from this distribution will be exactly captured by draws from the current distribution. This reflects the marginal treatment comparability assumption discussed in the previous section. In this case, causal associations derived from data may exhibit externally valid predictions to other periods of time.

In contrast, in very distant time periods, society may look completely different from today, such that conclusions drawn using data from the recent past have almost no relevance. It is hard to say at what point technological advances and other changes may create such a society but considering the rapid advancement of technology over the previous century, with inventions such as the internet, mobile phones, and television, it is likely we have very little understanding of life in 100 years' time. Similarly, the impact of climate extremes outside our known distributions could be hard to estimate. Sea-level rise provides a key example: the societal effects of permanent changes to the coast line are impossible to know based on existing data, as these changes have not yet occurred. In these cases, mechanistic models may provide the only possible modeling solution to demonstrate impacts. Agent-based models provide an example of such an approach–the emergent results of these models are typically not predictive of existing phenomena and parameters within these models are not calibrated to data but are based on estimations, expert elicitation or external studies.

For all other cases, hybrid models provide the best approach. Hybrid models may take many different forms but are essentially driven by two core components: structure within the model that elucidates some of the core pathways (the mechanistic component) and parameters that are determined at least partially by fits to data (statistical component). The mechanistic approach provides an advantage when considering out-of-sample future changes. If mechanisms driving a relationship between climate and an outcome are clear, and the linkage between these factors and other variables within the system are also somewhat clarified, then simulations outside of the range of experienced time and climate have more robust validity. The statistical approach provides an advantage by allowing these models to still be grounded in empirical data. This allows for the refutation of some of the previously discussed criticism of climate-impacts modeling: that parameters within models are essentially based on guesswork.

1.3.1 Examples of hybrid modeling approaches

Integrated Assessment Models, using damage functions estimated statistically from empirical data, are one example of a hybrid approach, with the integrated assessment model providing some mechanistic underpinning. However, in this section I shall consider explicitly models of a lower order, which attempt to establish predictions of outcome Y as climate variable C changes. In these cases, running a simple regression as in Equation 1.1 is defined as a primarily statistical approach. Predictions of the future effects based on such a regression may have limited external validity as climate variables change and/or society changes.

The next step is to include an interaction term between climate and covariates such as baseline temperature, population density and GDP. This approach is discussed in more detail in Chapter 4 and also briefly in Section 1.3. Estimating these interactions within the statistical model allows for the development of a more robust prediction

model. This approach is still mostly statistical; however, the interaction terms may be used to characterize a mechanistic model. For instance if we have know the effect of mortality conditional on GDP, and we know the effect of climate on GDP, we may be able to capture more complex feedbacks between climate, mortality and GDP, which in essence applies mechanistic structure across impact sectors. For a nice example of mechanistic modeling approaches that combine multiple sectors, see recent work by Ngonghala [94, 95].

For some sectors, mechanistic models are fundamental to the discipline. Epidemiologists predicting infectious disease incidence use a suite of models that capture the underlying process of disease transmission [2], resulting in predictions of incidence and prevalence. These models have a long history of generating successful predictions. In many cases demographic factors that may be expected to shift over time are already explicitly incorporated in these models–for instance population sizes and birth rates–making them a robust framework for starting to think about the future. In terms of the climate, dependence on climatic variables can be added to models depending on theoretical understandings of how climate changes might influence the processes at hand. For vector-borne diseases, as an example, climate may impact features such as vector biting rate, vector survival probabilities, egg survival probabilities and others, all of which can be incorporated into a mechanistic model [90],. The statistical element comes from fitting such a model to datasets on incidence, for example, of malaria or dengue cases. Techniques such as Bayesian MCMC allow for parameters to be fitted, such as the dependence of biting rate on temperature and also allow for the incorporation of prior knowledge from laboratory or other studies regarding the range that these parameters might take. Once such a model is parameterized it may be run into the future using a forced climate. Concerns over nonlinear responses at climate extremes may still apply to such models, but an explicit understanding

of how and where climate interacts with the system allows for better assessments of whether such responses might occur.

In Chapter 2, I apply a hybrid statistical-mechanistic approach to model changes to HIV prevalence given climate change. In this case I start with the regression technique used in Equation 1.1 to look at (relatively, i.e.six-year) short-run changes to climate and the associated effect on HIV prevalence. Over many decades HIV exhibits clear dynamics such that a more nuanced modeling approach is required for predictions at the decadal time scale. In this case I use my short-run results to parametrize a mechanistic model based on prior work suggesting the climate may drive changes to the proportion of individuals in the high-risk group for HIV and use this model to generate predictions for the future.

In Chapter 3, I present another hybrid statistical-mechanistic approach for generating climate change predictions of infectious diseases, this time for directly-transmitted immunizing infections. This approach builds off a mechanistic model of these diseases to generate a novel dataset of transmission rates which can then be used as the dependent variable in a panel regression such as Equation 1.1. Laboratory experiments have shown that climate conditions can affect the suspension of disease-carrying droplets, for certain airborne diseases, and as such, directly alter these transmission rates. The alternative to such an approach would be to regress climate on cases of the disease directly, but converting to transmission rates allows the mechanistic model to be used as the powerhouse behind climate-driven projections of the disease. Furthermore, the direct incorporation of population and birth rates in the model allows for alternate scenarios to be developed to consider future effects.

It should be noted that in this particular case, the nonlinear dynamics of the underlying system potentially bias even within-sample predictions if they are not accurately accounted for. Although complex systems are not included in Figure 1.2,

these might present another axis on such a plot. In these cases the size of the coefficient on C is dependent on the state of the system and must be appropriately controlled for.

The hybrid statistical-mechanistic modeling approaches developed in this thesis are discussed further in their respective chapters. It should be noted that one further advantage of semi-mechanistic models is that once the underlying systems have been characterized, policy interventions can also be considered. Implications for policy represent a third and often missing piece of climate-impacts studies. To some extent policy interventions are addressed in the chapters here, but I also consider in the conclusion how future work should take these into account.

1.4 Outline of the thesis

This thesis investigates the effect of climate change for three separate outcomes following the methods discussed in Section 1.2. The first two studies look at two potential health outcomes of climate change. Health effects may be *direct* or *indirect*, wherein direct effects occur through exposure to climate and indirect effects occur when climate alters the broader socioenvironmental context within which human beings interact with resulting negative impacts for health.

The first empirical study explores an indirect effect: the implication of climate change for HIV prevalence. The results build on prior work that suggests that behaviors associated with an increase risk of HIV exposure increase during periods of worsening environmental conditions. These behaviors include short-run migration and engaging in transactional sex markets. Using a dataset of over 400,000 individuals I find evidence that such behaviors do increase during warmer periods and furthermore HIV prevalence also increases. In order to predict future effects of climate change on this outcome, I develop a mechanistic model of HIV transmission

that I fit to country-level prevalence data. My results suggest that warming, coupled with population growth, could significantly increase incidence of HIV over the next 30 years if appropriate measures are not taken. I investigate the potential to curb this change in incidence by improving the distribution of antiretroviral drugs that have been shown to curb transmission. By running simulations over different antiretroviral coverage scenarios I find that changes in coverage, if enacted immediately, can completely remove the potential climate driven burden.

The second empirical study explores a direct effect: the effect on transmission of directly-transmitted immunizing infectious diseases. Such infections include measles, mumps, rubella, scarlet fever and pertussis. I study the association between humidity and transmission of varicella using data from Mexico. I find that drier conditions increase the transmission of varicella, a result that is supported by similar work on influenza, another airborne disease. Furthermore, the semi-mechanistic modeling approach I develop to look for these associations can be used to generate predictions under a forced climate. I find that predicted drying in Mexico towards the end of the century, according to climate change models, causes a shift in the timing of cases. Such shifts in timing may have important implications for public health officials considering control efforts for these diseases.

The final empirical study looks at the effect of climate change on labor supply in urban areas using an employment dataset with over 15 million observations. I find that hot days decrease time spent at work by over an hour for certain high-risk industries. Furthermore, I find that in cities in which urban heat islands exacerbate experienced temperatures, the labor supply in all industries appears to be affected. Using a simple model I estimate the coupled effect of urban heat island and climate change on labor supply until the end of the century. The changes in labor supply are monetized using city-level wage data and found to cost approximately $15bn 2013 US dollars cumulatively for the worst hit cities. City-level cooling measures may

be able to mitigate these losses. I find that the most cost-effective measure is cool roofs wherein the albedo of roofs is increased to reflect sunlight. Implementing this strategy in Rio de Janeiro, where the urban heat island effect is particularly strong, is a cost-effective solution to mitigate negative climate impacts on labor supply.

Chapter 2

Climate change drives modeled HIV prevalence

This work has been presented at Population Association of America Annual Meeting (2017), American Geophysical Union Annual Meeting (2016) and the Interdisciplinary PhD Workshop in Sustainable Development, Columbia University (2016).

Abstract

Long term changes in temperature can negatively affect human livelihoods with resulting implications for health outcomes. Here I leverage a dataset of over 400,000 individuals across 25 countries in Sub-Saharan Africa, coupled with high resolution climate data for the region, to test the effect of long run temperature changes on HIV prevalence. I find that warmer periods are linked with an increase in HIV prevalence. Both economic and behavioral changes likely drive this linkage: I find evidence that male migration and sex-market use increases with higher temperatures. In order to test the potential effect of climate change on HIV prevalence, I develop a simple model of HIV spread for each country in the dataset and fit the model to existing HIV prevalence data using Bayesian techniques. I couple my estimated results from the survey data with this mechanistic model to find the predicted changes in HIV prevalence for each country under all RCP warming scenarios. These changes in prevalence range from 0.22 to 4.39 percentage points by 2050 dependent on country-level factors such as existing HIV prevalence, antiretroviral coverage and the extent of future warming. In terms of incidence, the model suggests that warming under an RCP8.5 scenario could cause 15.44 million additional cases of HIV/AIDs by 2050 under current population growth rates.

2.1 Introduction

There are multiple pathways through which climate, and climate change, affect human health. Direct effects include increases in disease transmission, pulmonary and cardiovascular problems due to heat stress, and fatalities from extreme events such as flooding and wildfires [9, 30, 67]. Indirect effects occur when climate drives changes in the broader socio-environmental context within which human beings interact. For instance, climate-driven agricultural losses may cause households to migrate to urban areas where they can contract new diseases; individuals facing agricultural losses may be at increased risk of suicide; or climate-driven local violence can result in fatalities [82, 20, 64]. While short-run climatic changes drive these direct effects, indirect effects, or processes, may result from temperature shocks that occurred several years previously, or from multiple years of warming [27, 21]. A better understanding of these processes will lead to an improved characterization of future damages from climate change, and also to more effective policy interventions aimed at minimizing negative health outcomes.

Non-governmental organizations have for a while proposed a link between changing climates and HIV [13, 34, 50, 137]. A joint report by UNEP and UNAIDS, released in 2008, suggested multiple factors could link the two phenomena [137]. Of primary concern is that climate change may increase food insecurity, a well-studied determinant of changing HIV prevalence [48, 140, 125, 88, 136, 104]. Diminished food security is both a cause and a consequence of changes to HIV status [3]. For instance, the loss of productive household members to HIV/AIDS exacerbated food shortages in South Africa [25]. In contrast, a study in Botswana and Swaziland found that women reporting insufficient food in the previous 12 months had a 70% increased odd of engaging in unprotected sex, increasing HIV risk [140]. Furthermore, women suffering from food insecurity may engage in transactional sex, exchanging sex for food and gifts. These partnerships may increase the risk of contracting HIV, espe-

cially if a premium is placed on limiting prophylaxis use [33, 116]. Young women are particularly at risk of new infections if age gaps in partnerships are large [73, 100].

Climate-driven changes to food security are also linked with changing migration patterns [137]. Climate shocks can increase both short term migration and lead to permanent moves [12, 51, 52]. Migrants may be at a higher risk of contracting HIV if they engage in novel partnerships in different locations [14] or visit sex markets [68]. Individuals living in rural locations are expected to be particularly affected, as agricultural productivity decreases with rising temperatures[76, 120].

Despite the known link with food security, the relationship between climatic changes and HIV has received relatively little treatment in quantitative academic research [132]. Only one other study exists that attempts to quantify such a relationship: Burke, Gong and Jones found that a single rainfall shock, within a 10 year period, is associated with an 11% increase in HIV rates in Sub-Saharan Africa [18]. Data limitations may have prevented further exploration of the topic. This chapter takes advantage of the recent release of additional rounds of the Demographic and Health Survey data with HIV testing and GPS-located survey sites. The availability of this data has created a novel opportunity to look at long-term trends in HIV prevalence and environmental correlates.

I first establish the link between warming years and changes in HIV prevalence, investigating sub-groups that are particularly vulnerable. Then I explore the effect of temperature on several risks factors for HIV, to the extent it is possible within the data. Finally, I use a dynamic model to capture the interplay between temperature and HIV and generate predictions for future prevalence changes under alternate warming scenarios. The model is fit to data on HIV prevalence over time and the climate-dependence is parameterized based on the statistical results. In addition, this modeling approach allows for the introduction of a policy change in antiretroviral coverage, such that a possible mitigation solution can be analyzed.

2.2 Data

Survey data comes from the DHS. All surveys in Sub-Saharan Africa that include HIV testing and GPS-tagged survey locations for both men and women are included. Using multiple survey rounds enables the construction of a pseudo-panel to look at the effect of changing climates on HIV prevalence over time in a fixed location. In total, the dataset covers 400,000 individuals across 25 countries (Appendix Figures A.2, A.3).

Temperature data comes from the Global Meterological Forcing Dataset (GMFD) [99], at a resolution of 0.25*025 degrees latitude and longitude. Monthly temperature data is averaged across the six years prior to each survey (six years being the average length of time between survey rounds), and matched to individuals using GPS coordinates (Appendix Figure A.2). Temperature projection data is taken from CMIP5[5] using the multi-model mean at a resolution of 1*1 degree latitude and longitude. Data on national-level HIV prevalence comes from the World Bank [8] as does data on antiretroviral coverage and population.

2.3 Methods

Temperature data was matched with survey data and accompanying serological data in order to estimate the following linear probability model:

$$HIV_{igmy} = \beta f(T_{gmy}) + \alpha_g + \gamma_{c*y} + \varepsilon_{igmy} \tag{2.1}$$

The outcome HIV_{igmy} is a binary variable for whether an individual i tests positive for HIV in month m of year y in a location defined by a grid-cell g. The dataset is a repeated cross-section, not a panel, so the same individuals are not observed over time; however, I controlled for regional variation in HIV prevalence by including a

region-specific grid-cell fixed effect α_g. Grid-cells are defined by the climate dataset as a 0.25*0.25 degree square (approximately 25km-by-25km at the equator). Trends in HIV prevalence are controlled for using country-by-year fixed effects γ_{c*y}. Temperature $\beta f(T_{gmy})$ is measured as average maximum monthly temperature over a six-year lagged time period starting from month m and year y at location g. Six years was chosen because that is the average length of time between the two rounds of surveys that take places in each country. Using a longer period of data would mean that clusters would experience overlapping treatment. Error terms are specified as ε_{igmy}. I clustered standard errors at the country-by-year level to account for correlations in HIV prevalence across the country. The results are also robust to two-way clustering at the country and year level separately (Table A.1). I included HIV survey weights, as specified in the data, in all regressions.

Estimates for the effect of temperature on the potential causal pathways linking climate to HIV were found using:

$$Y_{igmy} = \sum_b \beta^b T_{gy}^b + \alpha_g + \gamma_{c*y} + \varepsilon_{igmy} \tag{2.2}$$

where Y_{igmy} is the outcome of interest for individual i as surveyed on month m, year y and location g. T_{gy}^b is the number of months in grid-cell g and year y where average maximum monthly temperatures fall into bin b. Bins are of $3°C$ in width. Other variables remain the same as above.

Mechanistic model

A simple compartmental model was used to generate HIV predictions. The model was adapted from [139] to include temperature. Further justification for this modeling choice is given in Appendix A. The model is governed by a series of differential

equations:

$$\frac{dS_H}{dt} = b * p_H * N - \lambda_H * S_H - m * S_H \tag{2.3}$$

$$\frac{dI_H}{dt} = \lambda_H * S_H - \gamma * I_H - m * I_H \tag{2.4}$$

$$\frac{dA_H}{dt} = \gamma * I_H - \mu * A_H - m * A_H \tag{2.5}$$

$$\frac{dS_L}{dt} = b * (1 - p_H) * N - \lambda_L * S_L - m * S_L \tag{2.6}$$

$$\frac{dI_L}{dt} = \lambda_L * S_L - \gamma * I_L - m * I_L \tag{2.7}$$

$$\frac{dA_L}{dt} = \gamma * I_L - \mu * A_L - m * A_L \tag{2.8}$$

At any moment, the state of the system is given by S, I and A where S is the number of susceptible individuals (the population which has never contracted HIV/AIDS); I is the number of individuals infected with HIV; and A is the number of individuals with AIDS - the advanced state of the disease where mortality rates are high. The H and L subscripts refer to the high-risk and low-risk groups respectively.The proportion of individuals in the high-risk group is given by p_H and is assumed to be constant across time. The background mortality rate for the whole population is given by m and the supply of new susceptibles is governed by b. μ is the AIDS-specific mortality rate and γ is the rate at which the HIV infection develops into AIDS. Appendix Figure A.6 represents this set of differential equations as a diagram.

Derived parameters are given by:

$$\lambda_H = c_H * \beta(\alpha * k) * q \tag{2.9}$$

$$\lambda_L = c_L * \beta(\alpha * k) * q \tag{2.10}$$

$$g_H = \frac{c_H * S_H + I_H}{c_H * S_H + I_H + c_L * S_L + I_L} \tag{2.11}$$

$$g_L = 1 - g_H \tag{2.12}$$

$$q = \frac{g_H * I_H}{S_H + I_H} + \frac{g_L * I_L}{S_L + I_L} \tag{2.13}$$

where λ is the rate at which susceptible (uninfected) individuals become infected. It is a function of the partner change rate c (i.e the number of new sexual partnerships a year), q the probability a new sexual partner is HIV positive, and β the transmission rate per partnership. q depends on g_H and g_L, which are the probability that a sexual partner is a member of the high-risk or low-risk group respectively, a function of the state variables.

The model was adapted to include the impact of antiretroviral drugs on transmission. Antiretroviral drug use α dampens transmission by $\alpha * k$ where k is some constant to be determined. k represents the fact that antiretroviral distribution may not have a uniform effect on transmission for a variety of reasons such as wrong usage, incomplete efficacy and poor reporting of actual coverage.

There are five unknown parameters:

- p_H the proportion of the population in the high risk group.
- c_H the partner change rate of the high risk group.
- c_L the partner change rate of the low risk group.
- k the modulating effect of antiretroviral drugs
- b the rate of supply of new susceptibles.

Other parameters are fixed:

- γ=(1/9) is the rate of progression from HIV to AIDs that takes approximately 9 years.
- $\mu = 1$ is the mortality rates of AIDs.
- $m = (1/35)$ is the baseline mortality rate of the population.
- $\beta = 0.05$ is the transmission rate per relationship.

Temperature enters the model through changing the proportion of individuals in the high-risk group (p_H). Therefore a relationship between the degree of temperature change and p_H over some fixed time period needed to be established. This was done in two steps. First, for every country the temperature-HIV (prevalence) response is estimated. Next, the change in p_H which could produce such a change in prevalence, over the time period of the survey, was calculated. Appendix Figure A.9 illustrates this process .

2.4 Results

2.4.1 Current response of HIV to temperature changes

A linear probability model was estimated from the data, with HIV status as the dependent variable. The results for the whole sample suggest that a 1 degree Celsius change in average monthly temperature, over a six-year period, increases the probability that an individual tests positive for HIV by 0.00419 percentage points ($p = 0.024$). This means that a 1 degree Celsius change (in average temperature over this period) leads to a relative 7.4% increase in the proportion of the population with HIV, given that the average prevalence in the sample is 0.056. This result is robust to alternate specifications (Appendix Table A.1). Further analysis reveals the sub-populations particularly at risk. Figure 2.1 shows the results for different age groups. HIV prevalence increases with age, making the young particularly vulnerable

to new infections [100]. The effect of temperature on HIV prevalence is greatest for younger age groups and diminishes over time such that it is negligible by age 40.

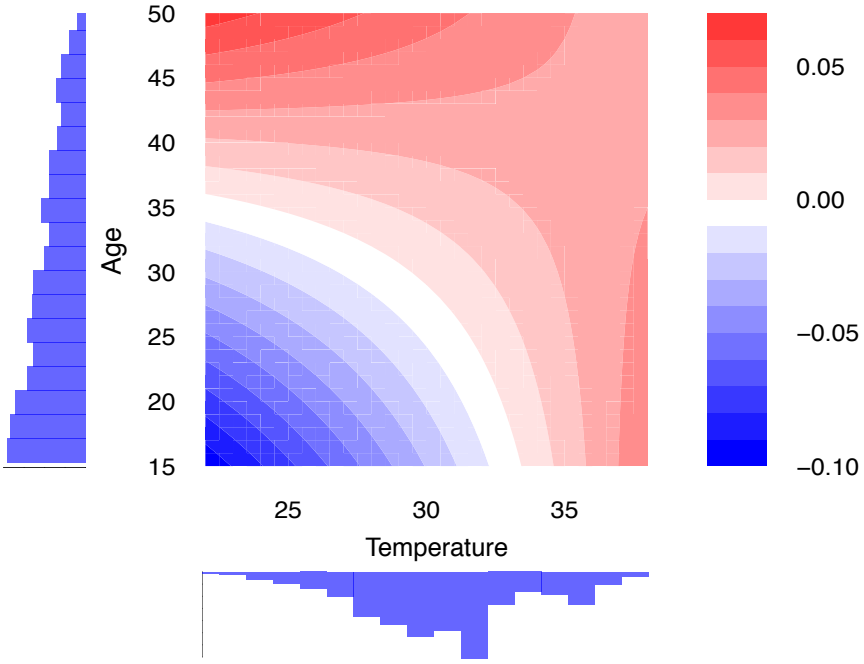


Figure 2.1: The effect of temperature changes on HIV prevalence for different ages, estimated using an interaction term between age and temperature. All terms within the model are significant ($p < 0.05$ **). Red regions show increasing HIV prevalence, while blue regions show decreasing prevalence. The results are centered such that mean age and temperature are set to zero and changes are interpreted relative to the mean. Histograms of the distribution of both age and temperature within the sample are shown by the respective axes.

Figure 2.2 shows the effect of temperature on HIV for different sub-populations: men, women, urban, rural and underlying prevalence. Sub-figure 2.2A shows the response over two different time scales: the short-run as an average over the previous three months and the long-run as an average over the previous six years. In the long run, hotter periods increase HIV prevalence. In the short-run there is a slight negative effect of temperature on HIV prevalence, though the effect is not significant for all sub-groups. This short-run effect may be driven by increased mortality amongst individuals with AIDS during hot periods. If heat-related mortality disproportionately

affects AIDS sufferers relative to non-HIV/AIDS then HIV prevalence will decrease during hot spells.

The long-term effect of warming on HIV is concentrated in rural areas (see Appendix Figure A.5 in addition). This supports the hypothesis that climate is related to HIV through economic damages in agriculturally-dependent regions that may result in food insecurity. Urban areas are more insulated from negative climate-economic effects, particularly for individuals who work in industries unrelated to agriculture.

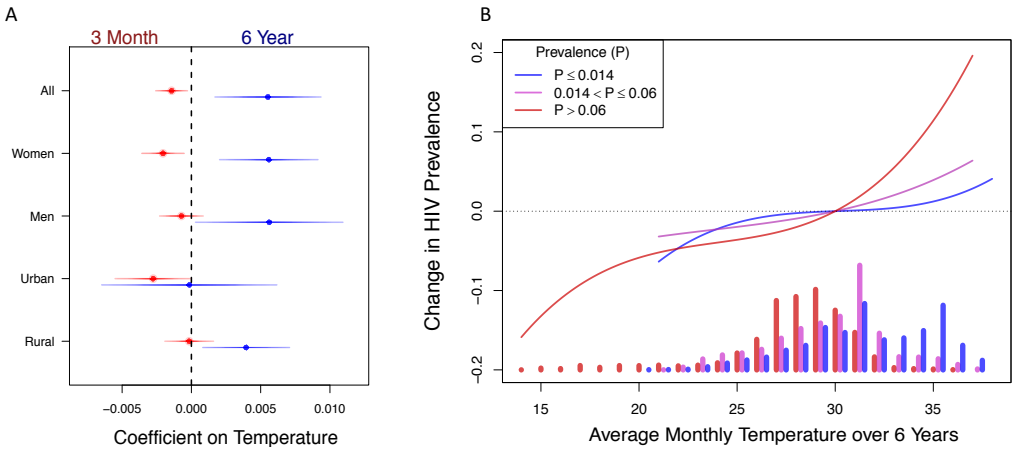


Figure 2.2: A) The relationships between temperature and HIV prevalence for different sub-populations showing both the long-term (6-year) and short-term (3-month) effect for women, men, urban and rural individuals. 95% confidence intervals on coefficients are shown. B) The effect of temperature on HIV prevalence for countries grouped by baseline (prior to survey) prevalence, estimated using third-order polynomials. F-tests of polynomial terms are all significant at the fifth percentile. A histogram of temperature distributions for each of these subsets of countries is shown.

The risk of contracting HIV in warmer periods is expected to be a function of current prevalence. Individuals in higher prevalence regions will be more likely to contract the disease if risk-taking behavior changes with temperature. Figure 2.2B evaluates the effect of temperature on HIV separately for three divisions of baseline prevalence. As expected, higher prevalence regions exhibit a steeper response to temperature, whilst low-prevalence regions exhibit a shallower response.

There are multiple possible explanations for the climate–HIV relationship. I used linear regression to test whether warmer periods change behaviors that are known to affect the risk of contracting HIV. Figure 2.3 shows these effects over two time scales: in the previous year (blue line) and over the previous six years (red line). A non-parametric binned temperature model was used that allowed for flexible testing of temperature–behavior relationships.

Women reported a decrease in the number of sexual partners when the previous year had been warmer; men, however, reported an increase in partners. Men also reported paying for sex more in warmer years. This suggests that mens additional

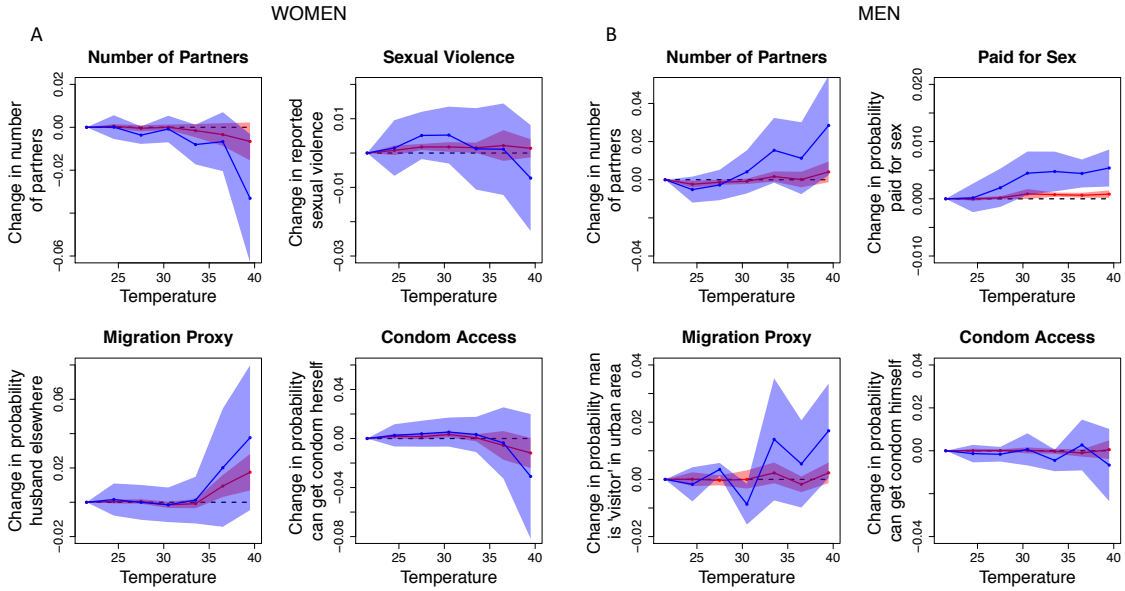


Figure 2.3: The effect of temperature changes on risk factors for HIV, estimated separately for women (A) and men (B). These results are estimated using a binned temperature regression where bins were constructed at 3°C intervals with temperatures below 20°C as the omitted bin. Monthly temperature was binned over the previous year (red) and the previous six years (blue). For both men and women, number of sexual partners (top left) and access to condoms (bottom right) was used as a dependent variable. Only female respondents were asked about sexual violence and only male respondents were asked about paying for sex (top right). The temperature effect on male migration was assessed by whether women reported their husband as "elsewhere" (A bottom left) and whether men were found to be visitors in urban locations (B bottom left). 95% confidence intervals are shown.

partners could be sex workers. Previous literature suggests that sex workers may be visited by migrant workers, so the effect of temperature on migration was tested [68]. The DHS does not explicitly ask about migration, so I looked at two variables that may be associated with population movement. First, I looked at whether women were more likely to report their husband as living elsewhere (other than in their home) when temperatures were higher. I found that in both the short-run and long-run, warming periods increased the probability that a woman reported her husband as living elsewhere. For men, I tested whether they are more likely to be registered as a visitor in response during warmer periods. I tested the effect only for men in urban areas, as the hypothesis is that warmer periods cause men to seek out labor in urban areas when agricultural jobs become limited. The results for this question are limited by the small sample size; however, the trend generally supports the hypothesis that men are increasingly found as visitors in urban areas in hot years.

I also tested the hypothesis that sexual violence increases at higher temperatures. I did not find this to be the case: in general, periods of warming do not increase the probability a woman reports experiencing sexual violence. Access to condoms also does not appear to be affected by temperature changes.

2.4.2 Modeling the future response of HIV prevalence to climate change

The results from analyzing the DHS data show that long-run temperature changes are causally linked to a change in community-level HIV prevalence. In order to assess the implications of these results for country-level changes in prevalence in response to future climate change, I constructed a simple compartmental model of HIV transmission and incorporated temperature dependence (Appendix Figure A.6). The model has several advantages over a more simplistic linear extrapolation approach: it accounts for the known non-linear dynamics of HIV infections over multiple decades as

well as allowing for manual adjustments to variables such as antiretroviral coverage to assess different response scenarios.

Individuals in the model are in one of three states: susceptible (i.e. not infected), infected (with HIV), or infected with HIV that has progressed into the symptomatic stage, AIDS. Furthermore, I assume that there are two sorts of individual in the population: high risk and low risk. This approach has been successful in other models of sexually transmitted disease [139, 60]. High-risk individuals are defined by having a much higher partner-change rate per year and can include younger age groups, migrants, sex-workers or women engaging in the transactional sex market. Temperature changes enter the model through a change in the proportion of the population in the high-risk category, as supported by the findings in Figure 2.3 as well as [18].

The dynamics of the model are governed by five parameters. These parameters are fitted using a Bayesian Monte-Carlo Markov Chain such that the model predictions match HIV prevalence data taken from the World Bank from 1990-2015. A separate model is fitted to each country, using only the countries included in the DHS survey rounds where antiretroviral data is also available ($n = 23$). The five estimated parameters are the proportion of high-risk individuals (prior to temperature changes), the partner-change rate of the high-risk group, the partner-change rate of the low-risk group, the rate of growth in new susceptibles, and the modulating effect of antiretroviral drugs (Appendix Table A.4). Antiretroviral usage enters the model by lowering the transmission rate [23].

Initially the model had no temperature dependence. Temperature was incorporated into the model by changing the number of individuals in the high-risk category. This reflects prior findings that temperature increases sex-market use and migration. The exact dependence of the number of high-risk individuals on temperature was calculated for each country using a simple algorithm based on the prior statistical results (Appendix Figure A.9 and Section 1.3). By including temperature dependence in the

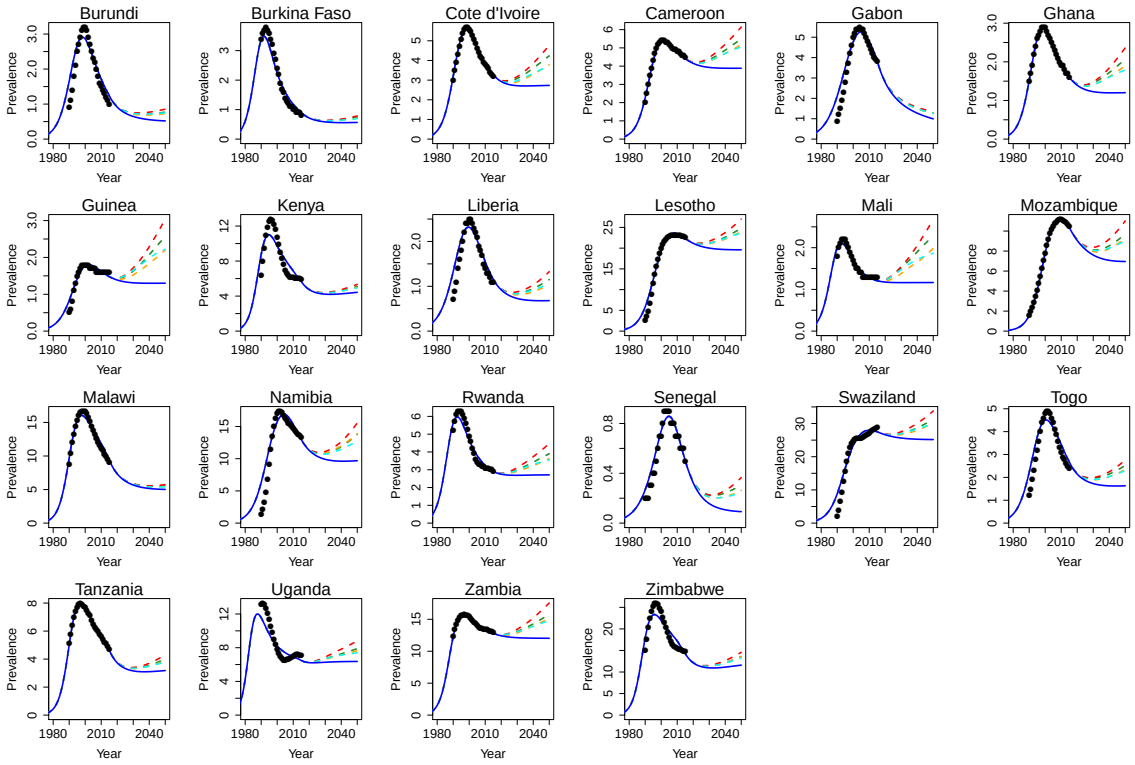


Figure 2.4: The model results for every each country where both HIV and antiretroviral data is available. Original HIV prevalence data is shown with black circles. The fitted model is shown with the blue line (without climate change). The model is was biased to fit best to most recent data, assuming that earlier prevalence data could be subject to underreporting. The dashed lines represent the predicted response of HIV for four warming scenarios out to 2050. RCP8.5 (red), RCP6.0 (orange), RCP4.5 (green) and RCP2.6 (cyan).

model, the future effect of climate change on HIV prevalence can be estimated. I generated these predictions for four different warming scenarios using the Intergovernmental Panel of Climate Change (IPCC) relative concentration pathways (RCP). The pathways are RCP2.6, RCP4.5, RCP6.0 and RCP8.5, with RCP2.6 representing firm action to reduce emissions and RCP8.5 representing the business-as-usual scenario where no action is taken.

Figure 2.4 shows the results for every country in the sample. The black points show the actual (World Bank) data on HIV prevalence for each country. The blue line shows the baseline model fit without considering climate forcing, and dashed lines show prevalence changes under the four climate scenarios. In all regions HIV prevalence

is expected to increase with climate change. The greatest change in prevalence is predicted in Swaziland at a 4.39 percentage point increase. The smallest change is in Burkina Faso at 0.22 percentage points. The variation in percentage point change is largely explained by prior prevalence and antiretroviral coverage. Changes in risk in countries with an existing high prevalence, such as Swaziland, result in a rapid spread of HIV amongst the new at-risk population. Countries such as Cameroon, with low antiretroviral coverage, experience a high relative increase in HIV prevalence.

Antiretroviral usage can lower the transmission of HIV by as much as 96% [23] if started soon after infection and used consistently. Increasing coverage could potentially be one policy option to mitigate the potential threat of climate change on HIV prevalence, with the additional benefit of alleviating the symptoms of HIV. I ran simulations using a range of potential antiretroviral coverage rates to calculate the level of coverage, if distributed immediately, that could minimize additional incidence from climate change out to 2050 under the RCP8.5 business-as-usual scenario. In these simulations I assumed that antiretroviral medication is taken within weeks of the initial infection and consistently such that the efficacy of antiretrovirals at reducing transmission is 96%.

The results of this analysis are shown in Table 2.1. On average, an increase in antiretroviral coverage of 9 percentage points offsets climate-driven changes to HIV prevalence. However, significant heterogeneity exists across countries in the optimal response (between a 1 and a 21 percentage point increase). Furthermore, differences in population numbers across countries can affect total numbers of infected individuals. Column 4 shows the potential number of cases reduced for a 1 percentage point increase in antiretroviral coverage in each country. The results suggest that in some countries (such as Tanzania, Mali and Guinea) increasing antiretroviral coverage could have a large effect on total number of cases. In other countries (such as Liberia,

Country	Current AR (%)	Recommended AR (%)	Cases Reduced
Burundi	54	61	8,489
Burkina Faso	55	60	11,200
Cote d'Ivoire	35	48	15,616
Cameroon	27	40	87,297
Gabon	58	61	3,529
Ghana	34	47	27,929
Guinea	29	40	176,813
Kenya	59	61	78,993
Liberia	24	43	1,831
Lesotho	42	46	105,411
Mali	28	46	167,583
Mozambique	53	58	117,769
Malawi	61	63	50,692
Namibia	69	70	7,536
Rwanda	79	80	7,264
Senegal	40	59	1,943
Sierra Leone	27	48	6,256
Swaziland	67	68	34,353
Togo	41	50	8,317
Tanzania	53	58	199,685
Uganda	57	62	56,787
Zambia	63	65	53,747
Zimbabwe	62	64	54,745

Table 2.1: Suggested increase in antiretroviral coverage to offset the potential effects of climate change. The second column shows current antiretroviral coverage (using 2015 World Bank Data - most recent available). The third column shows the recommended antiretroviral coverage to completely offset the total additional cases from climate change out to 2050. The fourth column shows the number of cases reduced per one percentage point change in antiretroviral coverage in each country.

Senegal and Gabon) increases in antiretroviral coverage may have less of an effect on total cases.

The total number of additional cases due to climate change between the present and 2050 was calculated under a variety of scenarios for population growth (Appendix Figure A.7). Countries within Sub-Saharan Africa are expected to experience continued high levels of population growth until the end of the century [46]. At the current population growth rate, 15.44 million additional cases are predicted amongst the 23 modeled countries, under the RCP8.5 warming scenario. The number of additional cases ranges from 8 million to 21 million as the population growth rate varies between 1.2% and 3.5%, reflecting the current range in growth rates amongst the modeled countries (Appendix Figure A.7).

2.5 Discussion

Predictions of socioeconomic outcomes of climate change are fundamentally uncertain. The well-documented uncertainty in climate models [49], coupled with unknowns in future population growth, economic development and technology, make generating predictions difficult. Adopting a hybrid statistical-mechanistic approach allows the incorporation of some changes in future variables: changes to antiretroviral coverage can be explicitly modeled, and different population growth scenarios are explored (Appendix Figure A.7). Other factors are not included: economic development or improvements to crop drought-resistance could reduce the number of individuals who enter the high-risk group under hotter conditions. It should be noted, however, that due to the non-linear dynamics of the disease system, even short-run changes in the number of high-risk individuals can have a resulting long-term change in HIV prevalence over multiple decades (Appendix Figure A.8).

Even if model predictions are uncertain, the potential cost of the realization of these outcomes, in terms of the detrimental effect to livelihood, may be very high. Increasing current antiretroviral coverage appears a win-win policy solution in terms of both alleviating the symptoms of the current HIV infected population and minimizing future cases. The results of the analysis on changes to antiretroviral coverage suggests that in certain countries, even a small change in percentage coverage, could have a significant effect on decreasing the total number of cases in the future.

The analysis here rests on two modeling approaches: a statistical model to estimate causal effects over the short-run and a dynamic model to generate long-term predictions. The statistical model prohibits the analysis of long-term changes by controlling for country-level trends over time; this is necessary to estimate a plausibly causal effect as trends could be driven by multiple factors. In contrast, country-level data on HIV prevalence is insufficient for causal estimation because trends over time are impossible to control for. However, this data can be used to parameterize a mechanistic model to predict changes to prevalence in the long-run. The mechanistic model used in this study is deliberately simplistic in order to be fit to data from different countries. Further details of the modeling choice are given in Appendix A.1.

Chapter 3

Dynamic response of airborne infections to climate change: predictions for varicella

This work has been presented at Ecology and Evolution of Infectious Diseases (2017), American Geophysical Union Annual Meeting (2017) and the Planetary Health Annual Meeting (2017).

Abstract

Characterizing how climate change will alter the burden of infectious diseases has clear public health implications. Despite our understanding of the transmission process for directly-transmitted immunizing infections, the impact of climate variables on these infections remains understudied. With co-authors[1], I develop a novel methodology for estimating the causal relationship between climate and directly-transmitted infections, which combines an epidemiological model of disease transmission with panel regression techniques. This method allows for progression beyond correlational approaches to studying the link between climate and infectious diseases. Further, the model can be used to generate semi-mechanistic projections of incidence across climate scenarios. I apply this method to 30 years of reported cases of varicella, a common airborne childhood infection, across 32 states in Mexico. I find significantly increased varicella transmission in drier conditions. I use this to map the potential changes in the

[1]This work is joint first-authored with Ayesha Mahmud, Harvard University and co-authored with Jessica Metcalf, Princeton University

magnitude and variability of varicella incidence in Mexico as a result of predicted changes in future climate conditions. The results indicate that the predicted decrease in relative humidity in Mexico towards the end of the century will increase incidence of varicella, all else equal, and that these changes in incidence will be non-uniform across the year.

3.1 Introduction

Understanding the link between climatic conditions and the transmission of infectious diseases has important implications for our ability to predict the timing and magnitude of outbreaks, as well as allowing better modeling of potential damages from climate change. Estimating such a relationship requires careful framing of the non-linearities inherent in disease dynamics, but also the interplay between climate forcing and other seasonal drivers of disease transmission, that may obscure their effects.

Much attention on the health impact of climate change has focused on non-communicable diseases [10], and vector and waterborne pathogens [117, 106, 105, 85, 141]. Directly-transmitted airborne pathogens, with the exception of influenza, have received relatively little attention despite evidence that climate may also shape the dynamics of these infections [78, 79, 124, 134]. Here, I leverage core understanding of the transmission dynamics of directly-transmitted pathogens [2, 38, 87], using as a case study, varicella, a common airborne childhood infection, in Mexico. My approach synthesizes two important methodological traditions: the use of panel regressions, a core tool in econometric analyses, with the use of mechanistic models, a founding element of models of infectious diseases.

Directly-transmitted, immunizing, childhood infections often exhibit a marked seasonal pattern in incidence, with regular annual or multiannual cycles of outbreaks [2, 87]. Both social and environmental factors may drive these dynamics as transmission depends on close contact (i.e. social settings) as well as the dispersal of

virus-carrying aerosols. A large body of literature has shown that the aggregation of children during school terms is a key driver of the observed seasonality and dynamics in countries with high rates of schooling [38, 11, 133]. The strong signature of social effects has perhaps diminished interest in the role of potential environmental drivers, yet known relationships between climate variables and viral-aerosol suspension suggest that climatic effects may also be important [79].

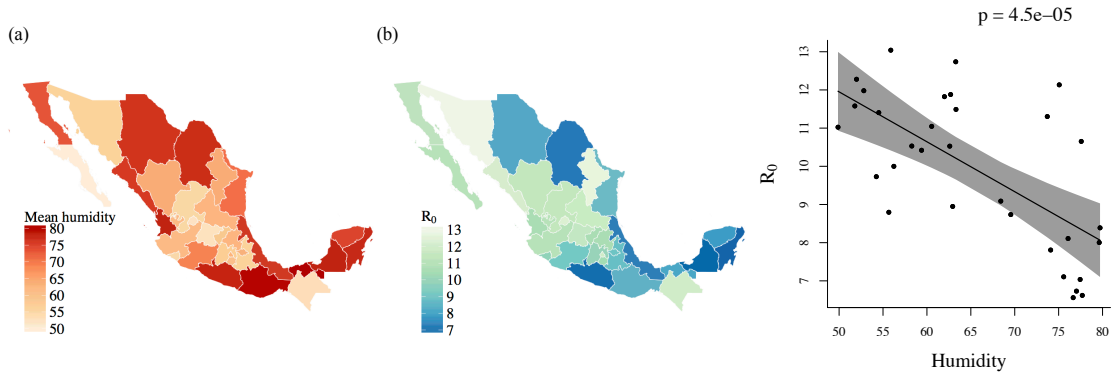


Figure 3.1: Relationship between average relative humidity and the estimated basic reproduction number, R_0. (a) Average humidity for each state in Mexico,1990-2007. (b) Estimated R_0 for each state in Mexico using age-structured case data from 1990-2007. For each state, R_0 was calculated as μ^{-1}/A, where μ = per capita birth rate and A = average age of infection (Anderson and May). (c) Association between R_0 and average humidity. Solid black line shows the estimated linear relationship; shaded region indicates 95% confidence interval; p-value noted at the top.

For directly-transmitted pathogens, the most conclusive evidence for an effect of environmental drivers on transmission is for the influenza virus. Experimental evidence from lab studies with guinea pigs [78, 79], integrated with population-based studies [134, 123, 124], have demonstrated that the transmission of the influenza virus, through respiratory droplets or aerosols, increases in drier and colder conditions; this observed relationship is likely due to the impact of humidity and temperature on the stability of the virus within aerosols and the stability of respiratory droplets in the air [79, 123]. Animal models allowing experimental tests of climate–disease relationships remain rare, yet inference into these relationships is of increasing importance. Cli-

matic conditions are predicted to change considerably in the future, and anticipating the associated changes in the burden of childhood and other infections will open the way to developing mitigation strategies.

Despite the availability of increasingly resolved data, methodological advances designed to quantify the links between climate and infectious diseases remain rare. Here, I develop a tractable and robust method to estimate the climate effects on transmission of immunizing childhood infections, in settings where social factors may also be at play. This modeling approach combines a mechanistic understanding of the dynamics of the infection, with panel regression methods to estimate the effect of humidity and temperature on transmission. The mechanistic model allows me to formally account for the non-linear dynamics of the disease, which is important because resulting population-level fluctuations in susceptibility might obscure the relationship between incidence and environmental factors. The panel regression model allows me to move beyond correlational approaches to studying the causal link between climate and infectious diseases by accounting for unobserved heterogeneities across states and over time. Further, I can generate semi-mechanistic projections of incidence that depend on climate conditions, to examine the impact of climate change.

I illustrate these methods using data on varicella (chickenpox), a common childhood infectious disease caused by primary infection with the varicella zoster virus (VZV) [43]. VZV is highly contagious and is transmitted through direct contact with skin lesions or via the respiratory route through the inhalation of aerosols containing the virus. Routine vaccination for varicella occurs in only a handful of industrialized countries, and the disease occurs worldwide with ongoing endemic transmission in many areas. In countries with endemic transmission, the morbidity burden of varicella can be high, resulting in considerable medical and non-medical costs [119]. Thus, improving our understanding of the drivers of varicella transmission and spread has important public health implications.

A conclusive link between environmental drivers and varicella transmission has not been shown previously, although a potential link has long been hypothesized because of observed differences in the epidemiology of VZV between temperate and tropical locations [42]. A few studies have documented a relationship between VZV incidence and climate conditions [77, 138], although these studies have been correlational. The availability of spatially-detailed temporal data on varicella incidence in Mexico (Appendix Figure B.1) offers a unique opportunity to explore the role of environmental drivers on transmission.

Mexico encompasses both tropical and temperate regions with temperatures ranging from 5°C to 35°C and relative humidity ranging from 20% to 90% (Appendix Figures B.2-B.4). Past work has shown that there is substantial variation in the mean age of infection of varicella across Mexico [80], which suggests spatial variation in transmission rates. Estimates of the basic reproduction number (the average number of secondary cases generated from the introduction of one infected individual into a fully susceptible population), suggests that transmission is typically higher in states that have lower average humidity (Figure 3.1).

3.2 Data

Incidence data for varicella was obtained from the Directión General de Empidemiología, Mexico. The dataset contains monthly counts for the number of reported cases of varicella between 1985 and 2016 for the 32 states of Mexico. Humidity and temperature data was extracted from the National Centers for Environmental Predictions (NCEP) North American Regional Reanalysis dataset, which provides long-term, high-resolution data for North America. Spatial averages of relative humidity and temperature data were constructed at a daily resolution for each Mexican state. I averaged the state-by-day data over biweekly time steps, to match the approximate

generation time for varicella. For climate change projections I used gridded humidity and temperature dataset from the fifth Coupled Model Intercomparison Project (CMIP5) under representative concentration pathway (RCP 8.5). RCP8.5 is considered a business-as-usual warming scenario wherein emissions continue to increase until 2100.

3.3 Methods

The novel estimation method combines an epidemiological model of disease transmission, the time series Susceptible-Infected-Recovered (TSIR) model, with panel regression techniques.

3.3.1 The TSIR model

The TSIR model is a stochastic, discrete time adaptation of the classic SIR model [38, 11], that can be used to reconstruct the unobserved susceptible population from data on reported cases and demographics. In the TSIR framework, the time series of the number infected and the number susceptible is described by a set of difference equations. For childhood infections, in the absence of vaccination, the number of infected individuals, I_t, tracks the births, B_t, and the number of susceptible individuals, S_t, which is unobserved in the data, is defined by:

$$S_{t+1} = S_t + B_t - I_t + u_t \tag{3.1}$$

where I_t and S_t are the numbers infected and susceptible at time, t, respectively; B_t is the number of births; and u_t is additive noise, with $E[u_t] = 0$. The time step for the TSIR model is the generation time of the infection which is approximately two weeks for varicella. To convert the observed monthly incidence into biweekly time steps I used linear interpolation to produce observations for 24 time points in the

year. I only observe a fraction of the actual number infected at each time step; thus I define the reported cases $Ir_t = \rho * I_t$ where ρ is the reporting rate.

The susceptible population at each time step can be written as $S_t = \bar{S} + Z_t$, where $\bar{S}$ is the mean number of susceptible individuals in the population, and Z_t is the unknown deviation around the mean number of susceptible individuals at each time step. I can rewrite the susceptible difference equation in terms of the deviations, Z_t, and iterate successively with an initial starting condition, Z_0. This yields:

$$\sum_{k=0}^{t-1} B_k = -Z_0 + 1/\rho \sum_{k=0}^{t-1} Ir_k + Z_t + u_t \tag{3.2}$$

Assuming u_t is small, Z_t can be estimated as the residuals from the locally varying regression of the cumulative number of births on the cumulative number of reported cases. The inverse of the slope of the regression line provides an estimate for the time-varying reporting rate, ρ. To allow for temporal variation in the reporting rate, I used cubic smoothing spline regression. Given estimates for ρ and Z_t, I can estimate the parameters of interest. At each time step, the expected number of infected cases is given by

$$E[I_{t+1}] = \frac{\beta_t I_t^{\alpha} S_t}{N_t} \tag{3.3}$$

where $E[I_{t+1}]$ is the expected number of infected individuals one infection generation time in the future; N_t is the total population size at time, t; β_t is the transmission rate; and α captures heterogeneities in mixing, not captured by the seasonality, and the effects of discretization. Log-linearization of Equation (3.3) yields:

$$ln(E[I_{t+1}]) = ln(\beta) + \alpha ln(I_t) + ln(\bar{S} + Z_t) - ln(N_t) \tag{3.4}$$

β_t is expressed as biweekly factors that capture the seasonal trend in the transmission rate, and captures the effect of both social and environmental drivers. The mean number of susceptible individuals, $\bar{S}$, can then be estimated using marginal profile

likelihoods from estimating equation (3.4) for a range of values of $\bar{S}$. The TSIR framework was used to reconstruct S_t for each state separately. An empirical estimate of the transmission rate, β_t, for each state for each time step, was calculated by rewriting Equation (3.3) in terms of the transmission rate, β, assuming $E[I_{t+1}] = I_{t+1}$:

$$\beta_t = \frac{I_{t+1} N_t}{I_t^{\alpha} S_t} \tag{3.5}$$

I constructed a panel of the empirical transmission rates by pooling the state-level time series into one dataset such that each observation is the transmission rate in state, s, at time, t.

3.3.2 Panel regression model

To estimate the effect of humidity and temperature on transmission I estimated the following linear regression model:

$$ln(\beta_{t,s}) = b_1 ln(H_{t,s}) + b_2 ln(T_{t,s}) + \gamma_{s*m} + \delta_{s*y} + \epsilon_{t,s} \tag{3.6}$$

where the dependent variable is log of the transmission rate of varicella at time t, in state, s, and the independent variables of interest are log humidity $H_{t,s}$ and log temperature $T_{t,s}$ at time, t. I included state-by-month fixed effects, γ_{s*m}, that control for known state-specific seasonality in disease transmission, such as state-level differences in school semester dates. State-by-year fixed effects, δ_{s*y}, control for state-level changes in transmission over time. These can include trends in migration, urbanization, changes in policies and other unobserved characteristics that vary year-to-year. Standard errors were clustered at the state level to account for correlations with the residuals at the state level over time. Panel regression estimates for b_1 and b_2 are the estimated effects of humidity and temperature, respectively, on transmission.

I assumed that the relationship between transmission and environmental drivers is log-linear; this functional form has been used in the past to model influenza transmission [134]. However, the results were robust to using an alternate specification where the relationship is described by an exponential functional form [124] (Appendix Figure B.6). In robustness checks I also tested for the effect of future humidity, as well as humidity from two weeks in the past, on transmission and, as expected, I only find an effect of contemporaneous humidity on transmission (Appendix Figure B.12).

3.3.3 Robustness check on simulated data

To test the robustness of my proposed method, I simulated varicella incidence using the SIR framework, assuming a linear and constant effect of humidity and temperature on transmission (Appendix Figure B.9), and examined the extent to which the estimation method could recover the underlying effects of humidity and temperature from the simulated incidence data. I simulated incidence by sampling from a negative binomial distribution at each time step:

$$I_{t+1} \sim NB(\frac{\beta_t I_t^\alpha S_t}{N_t}, I_t) \tag{3.7}$$

with mean, $\beta_t I_t^\alpha S_t/N_t$, and shape parameter I_t. I simulated S_t according to Equation (3.1), assuming $E[u_t] = 0$. I initiated each simulation with starting values I_0 (reported number of cases at the start of the time series, corrected for under reporting) and S_0 (from the TSIR susceptible reconstruction) and allowed the simulation to run for the length of the observation period (from 1985 to 2015). The transmission rate at each time step was calculate using Equation (3.7), with a constant value for b_1 (the effect of humidity) and b_2 (the effect of temperature). I included monthly trend in the transmission rate for all simulation scenarios, to capture the well-documented effect of school terms [80, 38]. I only included a yearly trend in the transmission rate

in one simulation scenario (described below). All simulation parameters, apart from b_1 and b_2, were obtained from the TSIR fit to reported cases, conditional on values of b_1 and b_2. I used a range of values for b_1, centered around the estimated effect from the model fit to reported cases. I fixed the value of b_2 at the estimated effect from the model fit to reported cases. I simulated incidence for each state for each value of b_1 assuming 1) constant population size and number of biweekly births over the entire time series (set at the mean of the observed values over the time period); 2) actual births and population size for the state; 3) assuming 10% to 30% under-reporting of incidence; and 4) assuming a time trend for the transmission rate that is independent from the effect of humidity. The time trend was obtained by fitting a cubic smoothing spline to the yearly trends from the TSIR fit to the reported cases for each state. For each simulation scenario, I compared the mean absolute error (absolute difference between estimated effect of humidity and the known effect of humidity used to simulate incidence, averaged across simulations done with different values of b_1) for the proposed estimation method, to applying the TSIR and panel models separately (described below). I ignored the transient dynamics at the start of the simulation by allowing a burn-in period of 300 time steps.

3.3.4 Comparison with other methods

I compared the results of the proposed estimation method applied to simulated incidence data, to results from applying the TSIR and panel models separately.

Comparison 1: TSIR model with environmental drivers

I used a modification of the TSIR framework, to estimate the effect of humidity and temperature on transmission of varicella separately for each state in Mexico. Typically in the TSIR framework, β is expressed as monthly or biweekly factors that capture the seasonal trend in the transmission rate. To incorporate the effect of climate variables,

β can be written as a function of humidity, H_t, and temperature, T_t:

$$ln(\beta_t) = b_{month} + b_{year} + b_1 * ln(H_t) + b_2 * ln(T_t) \tag{3.8}$$

where b_{month} captures monthly trends and b_{year} captures yearly trends in the transmission rate that are not explained by variations in humidity and temperature; as before, b_1 and b_2 captures the effects of humidity and temperature on transmission (and on incidence via its effect on transmission), respectively.

Equation (3.7) can be substituted in place of β in equation (3.4). The mean proportion of susceptible individuals, $\overline{S}$, can then be estimated using marginal profile likelihoods from estimating equation (3.4), with the full β specification, for a range of values of $\overline{S}$. Conditional on the estimated $\overline{S}$, I estimated α and each component of β using equation (3.4) with the full β specification. I use a quasi-poisson generalized linear model, which accounts for overdispersion, to estimate all parameters.

Since marginal profile likelihoods are used to estimate $\overline{S}$, fitting the model to all states simultaneously is intractable. The simulation results suggests that fitting the TSIR model to data on reported cases for each state separately produces a wide range of estimates of b_1 and b_2 (see Figure 3.2), and the bias in the estimates can be quite large depending on the functional form of the relationship between environmental drivers and transmission (Appendix Figure B.10).

Comparison 2: Panel regression on varicella incidence

I also compared results from the semi-mechanistic approach, to results from a panel regression of varicella incidence on environmental drivers. Unlike the proposed hybrid method, this method is purely statistical rather than mechanistic. I constructed a panel of the incidence data by pooling the state-level time series into one dataset such that each observation is the incidence in state,s, at time,t. I then applied standard

panel regression techniques to estimate the effect of humidity and temperature on varicella incidence. Specifically, I estimated the following linear regression model:

$$ln(I_{t+1,s}) = \alpha_1 ln(I_{t,s}) + \alpha_2 ln(I_{t-1,s}) + b_1 ln(H_{t,s}) + b_2 ln(T_{t,s}) + \lambda \mathbf{X}_{t,s} + \gamma_{s*m} + \delta_{s*y} + \epsilon_{t,s} \quad (3.9)$$

where the dependent variable, $I_{t+1,s}$, is log incidence of varicella at time, $t+1$, and the independent variables of interest are log humidity, $H_{t,s}$, and log temperature, $T_{t,s}$, at time, t. Climate variables are lagged by a two week period to reflect the generation time of the infection.

I included births and population in the vector of covariates, $\mathbf{X}_{t,s}$, since changes to population and births affect the total number of susceptibles, which constrains the size of possible cases. I also controlled for lagged incidence over the previous two time periods because of the autoregressive nature of disease transmission. In extensive simulations, I find that controlling for two lags of the dependent variable is the most appropriate lag structure in terms of minimizing bias in the coefficient (Appendix Figure B.11). It should be noted that using a lagged dependent variable in a panel model with fixed effects can lead to known biases [96]. However, these biases tend towards zero asymptotically as the number of time steps increases. Alternate model specifications with a differenced dependent variable, or without a lagged dependent variable, do not perform as well. I included state-by-month fixed effects, γ_{s*m} and state-by-year fixed effects, δ_{s*y}. Standard errors were clustered at the state level.

The linear panel regression approach allows identification of the effect of humidity and temperature on incidence from variations at the biweekly level within states. This method, however, does not fully account for the inherent non-linear dynamics of the infection and the change in the susceptible population over time which can

obscure the relationship between incidence and environmental variables (Appendix Figure B.11).

3.3.5 Projections

To explore the potential effect of climate change on incidence of varicella I used the simulation model (described above) combined with climate projection data from CMIP5. I compared incidence of varicella cases between 1985-2015 (simulated using historical climate) and 2070-2100 (simulated using projected climate changes), keeping constant all model inputs (population, births and seasonal trend in transmission rates), aside from the environmental drivers. Currently no population projections out to 2100 exist at the state-level in order for us to accurately change this variable in the model for different time periods. As such, the model represents a hypothetical illustration of the effect of climate change on varicella incidence, given the current demography of Mexico.

I applied the estimated effects of humidity and temperature on varicella transmission to the projection simulation. The main projection simulations were run without temperature changes, since I found no significant effect of temperature on varicella transmission. In order to ensure a fair comparison, I used CMIP5 data for humidity and temperature for both time periods, as opposed to using NARR data for 1985-2015. This ensures that I am not capturing the spurious effect of using humidity datasets with different data generation processes and resolutions.

Due to the stochastic nature of the model, I simulated incidence for both time periods for each state 500 times, with a burn-in period of 300 time steps for each simulation run. For each run I obtained mean values of incidence for each biweekly period of the year and computed the median across the simulation runs. As a robustness check, I simulated incidence assuming two different functional forms for the humidity-transmission relationship: a simple log-linear relationship [134] and an ex-

ponential relationship [124]. To examine the effect of predicted climate changes in Mexico, holding all else constant, I calculated the difference between the mean incidence during the two time periods (i.e. simulated incidence for 2070-2100 relative to 1985-2015) for each state.

3.4 Results

3.4.1 Causal Estimation

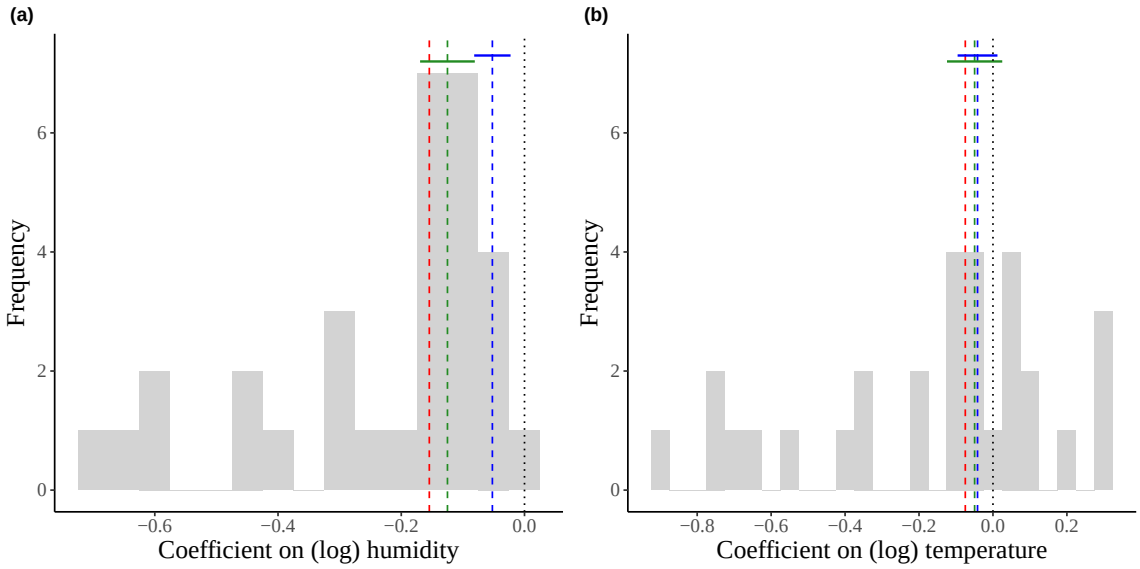


Figure 3.2: Estimated effect of (a) humidity and (b) temperature on varicella transmission, assuming a log-linear relationship between environmental drivers and transmission. The histogram shows the distribution of coefficients on log humidity and log temperature from fitting the TSIR model to each state separately. Only coefficients that were statistically significant at the 95% level were included. The red dashed line indicates the median of the TSIR coefficients on humidity (-0.154; s.d. 0.1997) and temperature (-0.075; s.d. 0.332). The vertical dashed green line indicates the estimated coefficient on log humidity (-0.125; clustered s.e. 0.0226) and log temperature (-0.050; clustered s.e. 0.0381) from the panel regression on reported cases. The horizontal solid green line indicates the 95% confidence interval. The vertical dashed blue line indicates the estimated coefficient on log humidity (-0.052, clustered s.e. 0.0151) and log temperature (-0.075; clustered s.e. 0.0275) from the combined estimation method (panel regression on empirical transmission rates). The horizontal solid blue line indicates the 95% confidence interval.

I fitted a time series Susceptible-Infected-Recovered (TSIR) model to data on reported cases of varicella, births, and population size. Empirical estimates of the transmission rate for each state and time step were obtained from the TSIR fit for each state. Panel regression with fixed effects was applied on the pooled dataset of transmission rate time series for each state, to estimate the effect of relative humidity and temperature on varicella transmission (see Methods).

Varicella incidence in Mexico is highly seasonal, with regular annual outbreaks that peak in April (Appendix Figure B.1). The peak in incidence coincides with the period of low humidity across the country (minimum average humidity occurs in April). Transmission for varicella is lowest during the summer months, which suggests that the aggregation of children in schools is a major driver of observed disease dynamics (Appendix Figure B.5). I control for variation driven by school terms by including month-by-state dummies in the panel regression.

I find that varicella transmission is significantly higher at lower humidities (Figure 3.2). The main result suggests that a 10% increase in humidity results in about a 1% decrease in transmission. I find no statistically significant effect of temperature on varicella transmission. The estimated negative association between humidity and transmission is robust to using an alternate functional form for the relationship between transmission and environmental drivers (Appendix Figure B.6) and also to estimating the model without including temperature (Appendix Figures B.6 and B.8). Further, using absolute humidity instead of relative humidity as a dependent variable in the panel regression, results in a similar sized effect on transmission, while temperature remains insignificant (Appendix Table B.1).

To verify that the proposed method can robustly disentangle social and environmental drivers of transmission, I simulated varicella incidence using the SIR framework, assuming a linear and constant effect of humidity and temperature on transmission (Appendix Figure B.9), and examined the extent to which the estimation

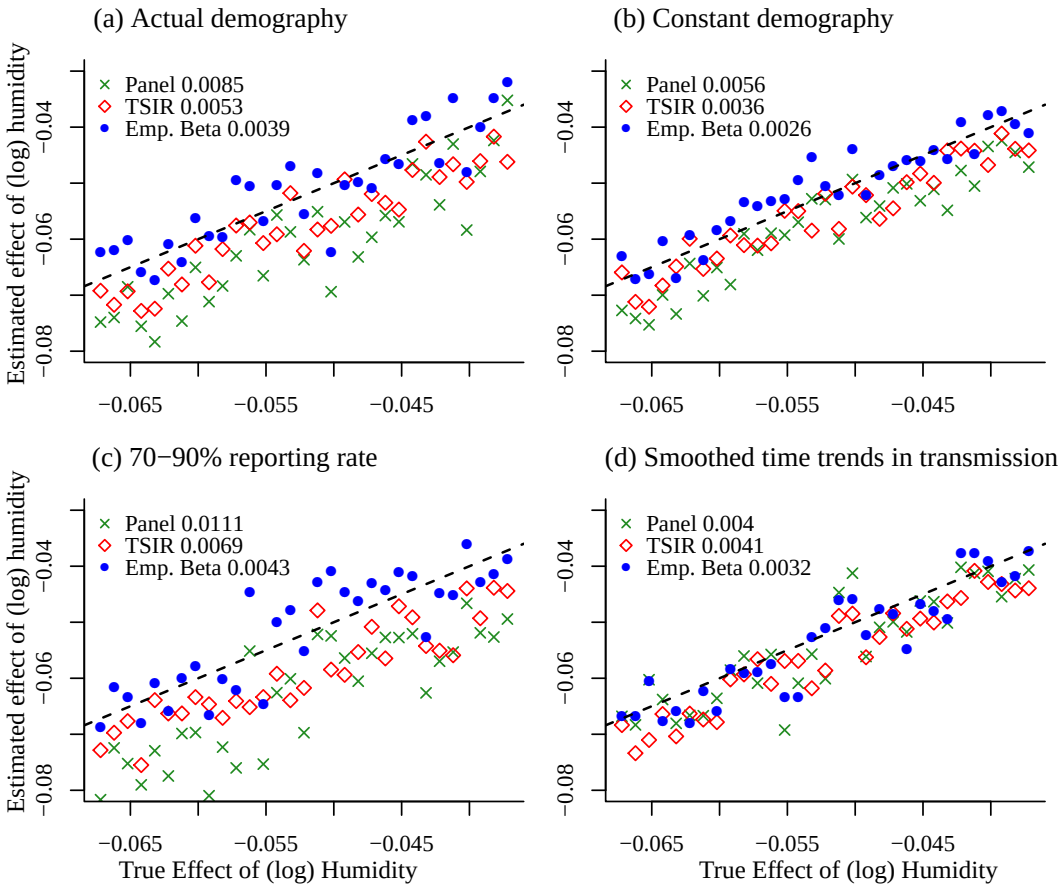


Figure 3.3: Robustness check of the estimation method on simulated incidence data. Varicella incidence for each state in Mexico from 1985-2016 was simulated using an SIR model under four scenarios: (a) using reported births and population size for each state; (b) using constant births and population size; (c) assuming that reporting rate for varicella increased linearly from 70% to 90% over the time period; and (d) allowing the transmission rate to have a time trend (the time trend was obtained by fitting a cubic smoothing spline to the yearly trends in the transmission rate from the TSIR fit). In each plot, the known effect of (log) humidity on transmission used to simulate incidence is plotted against the estimated effect of (log) humidity on transmission. The dashed black line is the x=y line. The green crosses represent the estimates from the panel regression on reported cases. The red diamonds represent the median of the estimates from TSIR model with environmental drivers fit to each state separately. The blue circles represent the estimates from the combined estimation method using panel regression on empirical transmission rates. The mean absolute error for each method is on the top left corner.

method could recover the underlying effects of humidity from the simulated incidence data. The simulation results suggest that across a range of scenarios, with different assumptions about the demographic variables, the temporal variance in the transmission rate, and the reporting rate over time, the method is able to identify the known effect of humidity on transmission. I evaluated the relative advantage of this combined approach compared to applying the TSIR and panel models separately (Figure 3.3). These results were robust to using an alternate specification for the relationship between climate and transmission and to applying the estimation method to incidence simulated without a temperature-transmission dependence (Appendix Figure B.10).

3.4.2 Projections

An advantage of the integrated approach is that I can use the mechanistic model to simulate the hypothetical impact of climate change on future incidence, conditional on the estimated transmission rate. I simulated incidence assuming no change in demographics and control efforts between current (1985-2015) and projection (2070-2100) time periods. For all Mexican states, relative humidity is expected to decrease by 2100 according to the fifth Coupled Model Intercomparison Project (CMIP5) under the representative concentration pathway (RCP 8.5). According to the simulation results, this drying is associated with an increase in incidence of varicella. Using two separate functional forms for the humidity-transmission dependence I estimate between 1000 and 2000 additional cases annually due to climate change for the years 2070-2100 compared to the current time period (Figure 3.4).

The simulations suggest that the effect of climate change on varicella incidence is likely to be non-uniform across the year. I find, on average, states will experience an increase in incidence in the summer months and a decrease in winter. Figure 3.5 shows the percentage increase in monthly number of cases of varicella for each state (states are ordered by mean expected humidity change). I find that states that

experience greater drying (between -3 to -4 percentage point change in humidity on average) also experience a more pronounced seasonal pattern with a loss of cases in winter months and an increase in summer months. States with a moderate amount of drying (-2 to -3 percentage point change in relative humidity) still exhibit a seasonal pattern, although increases in cases tend to occur later in the year. States with a smaller change in relative humidity do not appear to show a clear seasonal effect.

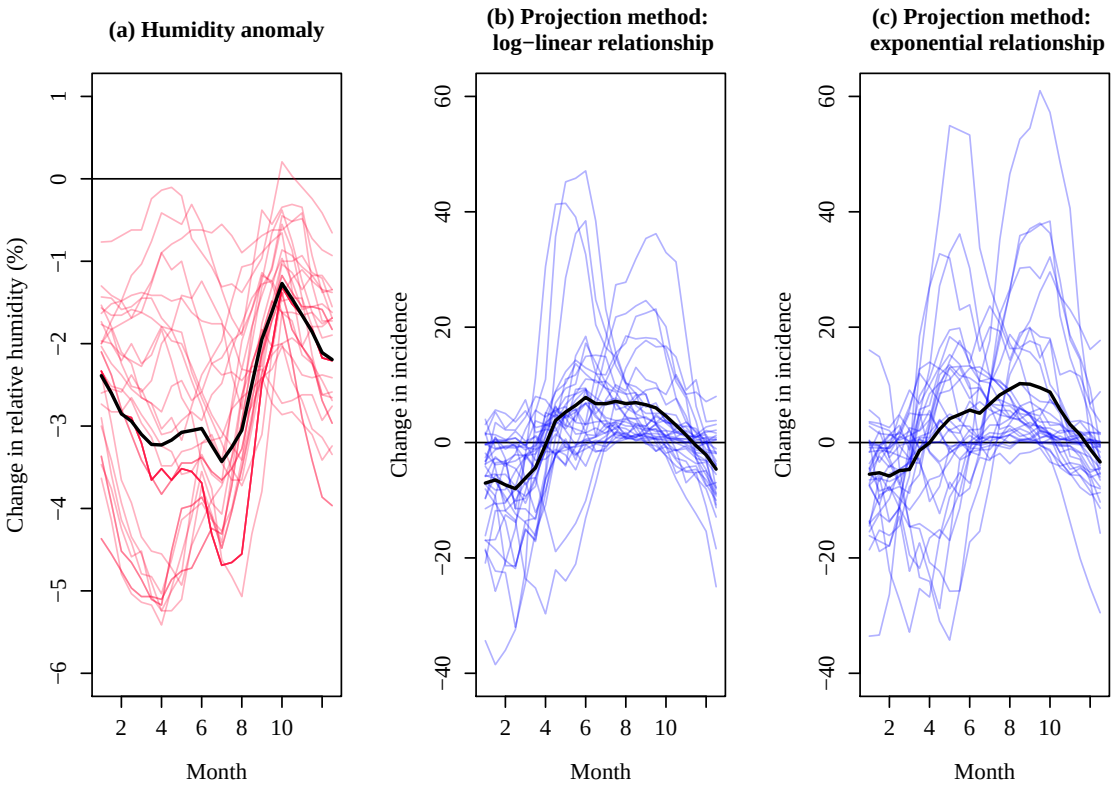


Figure 3.4: Projections of the impact of climate change on varicella incidence based on 500 simulation runs. Mean changes in relative humidity for each state by month are shown in (a). Mean changes in incidence for each state by month (i.e. change in simulated incidence for 2070-2100 relative to 1985-2015) are shown assuming two different functional forms for the humidity-transmission relationship (b) and (c). Projection method (b) assumes an underlying log-linear response of transmission to humidity and method (c) assumes an exponential relationship between transmission and humidity. The black line shows the mean change in incidence across states. The total change in annual incidence of varicella is 1025 cases using projection method (b) and 2015 cases using projection method (c).

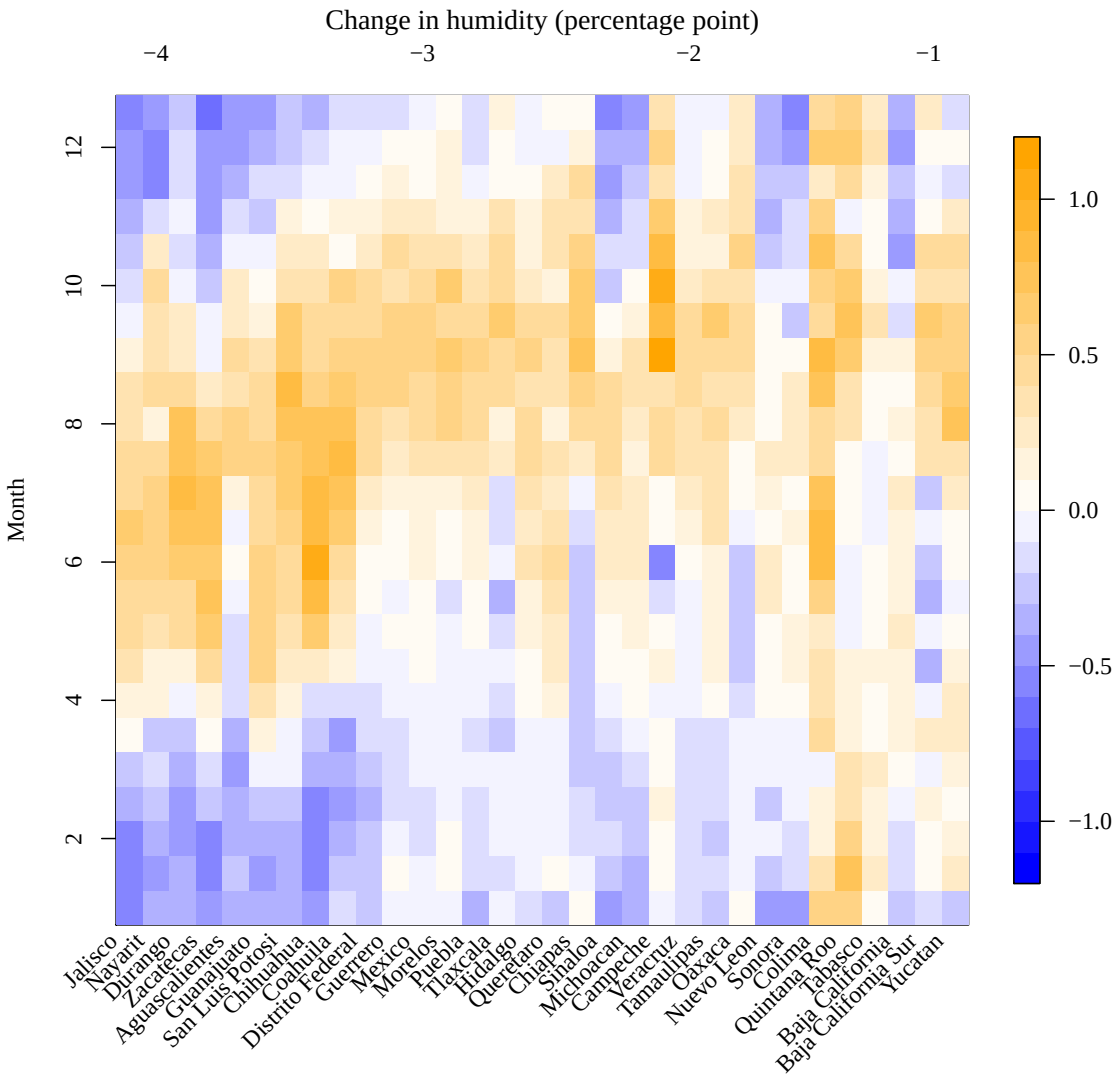


Figure 3.5: Percentage change in monthly incidence for each state for the years 2070 - 2100 relative to the years 1985-2015. Higher percentage changes are colored yellow and lower changes are colored blue. States are ordered by mean humidity change. States on the left experience the greatest change in average annual relative humidity (approximately 4 percentage point decrease). States on the right experience the smallest change (approximately 1 percentage point decrease). States with larger changes in relative humidity exhibit a marked seasonal pattern where cases increase in the summer and decrease in the winter. States with a moderate humidity change exhibit some gain in cases towards the late summer and early fall. Small changes in humidity do not result in a clear effect.

3.5 Discussion

The integrated model framing provides a robust and tractable approach to characterizing the impact of climate drivers on infectious disease transmission, accounting for social factors, and providing scope for simulating scenarios of future burden. Furthermore, this case study reveals a previously undescribed effect of humidity on varicella incidence. The results are in line with similar findings for influenza that has shown a negative relationship between humidity and transmission [79, 124, 134], and this framing opens the way to expanding this analysis to disentangle the relationship between climate and disease for other directly transmitted immunizing infections. This method may be particularly relevant for infections that lack an animal model, and for which direct experimentation is thus not an option.

Even with a clear mechanistic understanding of the link between climate and disease, forecasting changes in the burden of infection is complicated by the array of other important drivers that are likely to change over the future. In particular, state-level birth rates and population density are likely to change across Mexico, in ways that will affect varicella dynamics. Detailed projections of state level demography are not available, yet urbanization, economic development, and public health efforts to control varicella are all likely to change. Further, climate projections for changing humidity are deeply uncertain. In this illustration of the analysis, I used the multi-model mean, and did not account for the high variance across climate models [71], though such an analysis can be incorporated easily in future work.

The results indicate that changing humidity is likely to have a nonlinear effect on varicella transmission. Decreasing humidity does not uniformly increase cases but instead interacts with the typical seasonal pattern of varicella transmission to shift cases across time. This has important implications for policy makers from two standpoints. First, policy makers predicting disease incidence need to take into account that shifting climatologies may require planning for additional cases in months when these

cases would not typically occur. Second, policy makers involved in estimating damages from climate change need to include the non-linearity of disease dynamics when predicting outcomes. A simple linear treatment is insufficient to account for the complexity of the interaction. Expanding this analysis across other directly-transmitted airborne infections in different locations and climatologies is a key direction for future research, and an important step towards establishing the generality of patterns identified here. Extending the methods to encompass pathogens with slightly more complex signature of immunity, and which reflect stronger non-linear forcing (e.g., RSV [112]) is another important avenue.

The hybrid statistical-mechanistic approach developed here provides a tractable yet consistent method for probing the relationship between climate and disease for these pathogens. The ability to simulate the trajectories of communicable diseases under a changing climate provides a unique public health opportunity to prepare for the future. Furthermore, the grounding of the mechanistic model allows for different scenarios of both climatic changes and demographic changes to be explored. The provides a valuable tool for exploring potential risks associated with a changing climate.

Chapter 4

Implications of climate change for urban labor supply

Part of the work in this chapter has been presented at the American Economic Association Annual Meeting (2017).

Abstract

Multiple studies have linked economic output to environmental extremes. Changes to both labor supply and productivity are often cited as a potential driver of this linkage. In this chapter I explore the effect of temperature on labor supply in a specific context: urban areas in Brazil. Leveraging a high resolution multi-annual employment dataset, I find that hot days decrease labor by up to 80 minutes per day for individuals working in the highest-risk sectors. Using simple assumptions I estimate that the cumulative cost of this reduction could equal over $15 billion (2013 dollars) by 2100 in the hardest hit cities, as the coupled drivers of climate change and urban heat island reduce labor in urban environments. For Rio de Janeiro, for example, this corresponds to around 5% of city GDP. Interestingly, completely reducing the urban heat island effect fully mitigates these damages, even as the climate changes. City cooling measures are investigated – implementing cool pavements appears the most cost-effective strategy to reduce negative labor outcomes and combining cool roofs with cool pavements results in the maximum reduction in losses.

4.1 Introduction

There are multiple pathways through which ambient environmental conditions may affect both labor supply and labor productivity. Manual labor generates heat within the body that the body regulates through heat exchange with the environment. Warmer environmental temperatures reduce the efficiency of this heat exchange and temperatures in excess of 39°C can lead to heatstroke, a potentially fatal condition [70]. If physiological regulation fails to maintain body temperature individuals often alter their behavior to minimize risk. For instance, laborers exposed to high temperatures while undertaking physical activity may take more frequent breaks from work [81].

In addition to physical activity, mental cognition is also affected by prolonged exposure to high temperatures [144, 54]. Laboratory studies suggest that the ability to perform simple mental tasks and perceptual motor tasks deteriorates at around 30–33°C (measured in terms of Wet Bulb Globe Temperature (WBGT)) [114]. A meta-analysis of laboratory studies on temperature and performance found that different sorts of task are differentially affected by temperature. Mathematical tasks are affected by temperatures over 27°C (WBGT), whereas reasoning and learning tasks are more negatively affected by cold temperatures (<10C WBGT) [110].

Outside of the laboratory, panel studies from different industries have also been used to study labor productivity. Worker productivity in garment factories in India has been found to be negatively affected by temperature [128]. Interestingly, when LED lights were introduced to replace high-heat emitting fluorescent lighting, the negative temperature–productivity relationship was attenuated by 75% [1].

Labor supply is defined as the time a worker spends working and is thought of in terms of the extensive margin (the number of hours an individual works, given they supply work on a particular day) and the intensive margin (whether the worker works at all that day). Workers in adverse conditions may limit their supply of labor to minimize health risks. These adverse conditions are typically considered in

high-risk industries such as agriculture, where individuals are exposed to the outdoor environment. However, as temperatures increase, office workers may also be affected, especially if conditions are not regulated by cooling systems [122, 121].

Studies on manufacturing in India found mixed results for the impact of temperature on the intensive margin. Heat exposure affected absenteeism for workers in garment and rail mill industries; however, weaving workers were not affected. The nature of the wage contract may affect absenteeism. Salaried workers, who get paid even when absent, may have an incentive to stay home when conditions are poor. Workers who only get paid when they supply work (the case with weaving workers in India) may need to work even under extreme heat [128]. Research on time allocation in the US found that individuals reduce hours worked in "high exposure" industries and time spent outside when temperatures are hot [143]. In contrast, on rainy days individuals shift time allocation from leisure to labor [24].

Under extreme climate conditions, multiple changes within the socio-environmental system can affect an individual's labor supply. These changes to the system may be complex and non-linear. Blackouts due to a surge in energy demand from air-conditioners may cause whole industries to be shut down, for example. Urban areas may face specific threats such as changes to the functioning of transport systems. This further supports the necessity of context-specific analyses of climate change effects.

The existence of urban heat islands – the phenomena whereby urban areas attain, and at night maintain, hotter temperatures than nearby rural areas – may exacerbate the effect of hot days on labor supply in urban areas [74]. There are multiple causes of urban heat islands: manmade construction materials tend to be darker and therefore absorb more incoming radiation, lack of vegetation reduces moisture availability and therefore evaporative cooling; tall buildings limit air circulation; and energy usage in the city, for instance from air conditioners, can increase heating. The combination

of these factors can result in a dramatic temperature differential between rural and urban locations, with recorded maximums of +7°C [44].

The implications of urban heat islands for a warming world are still an active area of research. A key question is whether the temperature experienced in urban areas is a linear combination of the urban heat island effect and the ambient temperature (if urban development were not present). This is important when considering the implications of heat waves in urban areas. Model results suggest that heat waves may exacerbate the urban–rural differential, such that experienced temperature is even greater than the sum of these two components [74]. This response may also be a function of regional humidity [142] It is important to consider the additional effect of heat islands as a compounding factor in climate change damages in urban areas.

Interest in the relationship between labor supply and climate is driven by the linkage with economic output. Multiple studies have linked climate extremes to changes in economic output and production [26, 63, 29, 17]. Explaining the causal relationship between climate changes and macro-level economic changes is of importance for policy makers considering interventions to minimize damages.

4.1.1 Global Climate Prospectus

Accurate estimates of the social cost of carbon rely on projections of the future cost of climate change on all sectors of the economy for all regions in the world. The difficulty with constructing these estimates is that data limitations often restrict the assessment of climate driven damages to certain locations. One solution is to combine datasets from multiple regions, wherever data is available, and use interactions with fundamental, common characteristics across regions to generate a response functions that can be extrapolated to different time periods and locations.

The Global Climate Prospectus (GCP) is a collaborative research project with team members across several institutions which aims to improve estimates of the social

cost of carbon. The initial results of this chapter, i.e. the estimation of temperature effects on labor supply in Brazil, have been used in this global analysis for the labor sector of this project. The Brazil results are combined with eight other countries where labor supply data is available: USA, Nicaragua, Mexico, India, Guatemala, France, Spain and the UK [7]. Data on income, population density and average temperature are used for each sub-region available to look at how the effect of temperature on labor responds to these underlying characteristics. The result is a response function that is dependent on realized temperature, long-run mean temperature, income and population density. Using projected values for these variables allows for estimates of the future effect of climate change on labor supply.

The value of climate impacts studies is not just in estimating the social cost of carbon. Predictions of future damages can propel policy makers to take adaptive measures now to reduce potential impacts. In this chapter I initially estimate temperature effects on labor supply, then I look at the coupled effect of climate change and urban heat islands on labor supply. Finally I look at potential policies to reduce the urban heat island that could additionally mitigate negative consequences of climate change.

4.2 Data

The labor force data for Brazil comes from the Instituto Brasileiro de Geografia e Estatstica (IBGE). IBGE conducts multiple surveys. I use the Pesquisa Mensal de Emprego (PME) – a weekly survey taking place in six major cities. The survey is used by the government to track the status of the labor force in Brazil, including monthly unemployment rates and earnings. Data collection for the PME began in 1994 but the survey was significantly altered in 2002 into its current format, so only post-2002 data is used for this study. The spatial coverage of this dataset is limited

to six major cities whose total population represents approximately 13% of Brazil's 204 million people (see Appendix Figure C.1).

Survey data is collected on a weekly basis. Four weeks of data are collected for each month, meaning there is a total of 48 weeks of responses per year; this means that four weeks are not surveyed annually. The rationale for removing these four weeks is to create comparable employment statistics for each month. The weeks that are chosen to represent a month are those with the maximum number of days overlapping that month. The weeks that are dropped are the remainder.

The dataset is structured as a rotating panel, such that each respondent is interviewed eight times across the course of the survey. A respondent enters the survey on a particular week of the month and is surveyed again on that same week of the month for four consecutive months and then again for those same months one year later. For instance, someone may enter the survey on the third week in February 2000 and would then be surveyed again on the third week in March, April, May 2000 and the third week in February, March, April, May 2001. No individual id is provided in the survey, so I match individuals across time using household id, gender and birth date.

In order to make our dataset comparable to the other employment datasets used in the GCP study, the sample is restricted to all respondents in the work force i.e. those aged 15-65 who reported working more than zero hours in a particular week. In total there are 6.6 million observations.

Respondents are asked about how many hours they worked in the previous week where a week is defined as Saturday to Sunday. Temperature is matched at the city level by using cumulative temperature across the previous week. In the main specification, I take an agnostic approach with regard to whether respondents work on the week days or weekends, so temperatures for Saturday and Sunday are included. Holiday weeks are not included, however, in case they bias survey results.

The survey asks respondents about their primary source of employment. There are multiple detailed activity categories. Mining, agriculture and forestry are classified as primary sector employment; manufacturing, construction and utilities as secondary sector, and service, transport, retail and trade as tertiary sector. Agriculture, mining, construction and manufacturing are classified as high-risk employment categories; all other jobs are low risk.

Climate data in the main specification comes from the Berkeley Earth Surface Temperature (BEST) dataset. BEST data is available at a 1 degree resolution globally. Average of temperature is calculated within each city boundary [1]. Polynomial datasets are constructed by first calculating squared, cubed, etc. temperature data at the level of the gridded dataset and then averaging over space. Data for climate change projections comes from CMIP5 under a RCP8.5 warming scenario using the multi-model mean also at a 1 degree resolution. Further details of normalizing CMIP5 data with BEST data are given in section 5.2.

Gridded temperature datasets often remove the effect of urban heat islands as a potential source of bias [56]. Datasets such as Berkeley Earth make adjustments that may reduce the signal of heat islands. Typically, urban heat islands are studied using satellite-based measurements of surface temperature (e.g.[142, 66]), most commonly with the MODIS satellite. Data for calculating urban heat island effects comes from the Center for International Earth Science Information Network (CIESIN) [39] and is calculated for urban areas in the year 2013 using equivalent municipal boundaries. The data is based on eight-day composite MODIS land surface temperature data. The dataset provides both daytime and nighttime rural–urban temperature differentials. Only the daytime differential is used in this study, based on temperatures recorded at 1.30pm.

[1]For the GCP project, population-weighted temperature is used, however due to the small size of city boundaries, weighting across gridcells does not generally change temperature values at the city scale, as such all other temperature data in this study is area-weighted .

4.3 Methods

4.3.1 Temperature driven changes to labor supply

The effect of temperature on labor was estimated using the following model:

$$W_{it} = \sum_{k=1}^{k=4} \beta_k T_{ct}^k + X_i + \alpha_{c*m} + \gamma_{c*y} + \theta_w + \varepsilon_{it} \tag{4.1}$$

W_{it} represents the self-reported minutes worked by an individual i at time t. $\sum_{k=1}^{k=4} \beta_k T_{ct}^k$ is a fourth order polynomial function of temperature, where k is the order of the polynomial term and β_k is the fitted coefficient on temperature raised to the power of k. The main specification uses three fixed effects. α_{c*m} are month-by-city fixed effects controlling for variations across cities in monthly hours worked (Appendix Figure C.2). γ_{c*y} are city-by-year fixed effects, controlling for trends in working hours at the city level over time (Appendix Figure C.3) and θ_w are week level fixed effects. .X_i is a vector of individual level characteristics, in particular age, age^2 and household size. Standard errors were clustered at the city level.

Sub-groups were also analysed using the above specification but applying the model separately to individuals in high-risk employment, low-risk employment and the primary, secondary and tertiary sectors. City-level regressions were run to test for differential temperature responses across cities.

The GCP team uses an alternative specification with location-specific month-of-sample fixed effects. Results are shown using this specification for comparison with the GCP estimated global response.

4.3.2 Projecting the consequences of climate change

CMIP5 data is used to get projections of future temperature changes for each city in the dataset under the RCP8.5 level (Appendix Figure C.4). The CMIP5 data

sources gives monthly level temperatures. From this data, monthly level changes in temperature are calculated from 2013 until 2100, i.e. the 2013 January temperature is subtracted from future January temperatures to create a dataset of temperature changes for each location. Daily level temperatures are generated by randomly drawing a temperature from the distribution of historic daily temperatures in a location and then applying the degree change in mean warming calculated above. A dataset of daily temperature is thus generated, with the same standard deviation as historic temperature data, but shifts in mean temperature given by the CMIP5 projected changes.

The estimated response function for labor on temperature is applied to the projected temperatures in order to generate estimates for lost labor supply for an individual in each city, given climate change. The offset of this response function is adjusted such that the typical temperature for each day and city is set to zero. This normalization means that only deviations in labor supply from typical work hours are reported. This individual level estimate is coupled with population projections for each city. These projections are generated using a dataset of Brazil urban population numbers from 2013 to 2050 developed by the World Bank. Population growth rates are derived from the dataset and extrapolated to 2100 using a fitted polynomial bounded above 1. The projected growth rates are then applied to current estimates for urban population in each city provided by the IBGE. The data is summed at the annual level for ease of comparison between cities.

It is assumed that sectoral make-up of the city does not change over time such that the same proportion of individuals work in primary, secondary and tertiary sectors. Figure C.5 shows that sectoral make up of the cities studied has remained roughly equivalent over the last 10 years. Furthermore, the primary sector, which typically diminishes as a region develops, is already a very small percentage of the total work force for these urban areas ($<1\%$). Changes between the secondary and tertiary

sectors will not affect predicted losses in work time as individuals working in these two sectors exhibit a similar climate response.

Finally, an assessment is made of the potential cost of future reductions in labor supply. It is assumed that the value of labor is equivalent to the wage. Mean wages are calculated for each city, after removing outliers. The cumulative cost of climate driven reduction in 2013 wages is calculated from 2013 to 2100 without discounting.

4.3.3 Adjusting responses to account for urban heat islands

Urban heat islands can create elevated temperatures within urban areas. These elevated temperatures may exacerbate the negative consequences of climate change if the additional warming from urban heat islands remains as global temperatures increase. Furthermore, if urban heat islands effects are not accurately captured by temperature datasets, then estimates of climate impacts in urban areas may be biased.

It is difficult to measure the temperature that individuals actually experience ([142]: gridded datasets often make corrections to minimize the signal of urban areas on patterns of warming.[2] These adjustments are useful when assessing signatures of climate change, but they are problematic when accurate measurements of experienced temperature are required.[3] The relatively low resolution of current gridded temperature datasets, e.g. 1 degree * 1 degree latitude and longitude, means that temperatures are spatially averaged across urban and rural areas. This limits the recovery of temperature extremes in the small but densely populated urban cores.

The sum of these effects suggests that gridded temperature datasets may not accurately account for urban heat island effects. These datasets are downward biased such that the temperatures recorded may be lower than those actually experienced. Figure 4.1 shows average (maximum) temperature data from BEST in the year 2013

[2] See http://berkeleyearth.org/understanding-adjustments-temperature-data/

[3] Smart phone measurement of local temperatures may be one solution to this problem in the future.

over an area that covers the six cities used in the employment survey. The points in Figure 4.1 are the locations of the cities, with city names, and the numbers next to the points are the mean, 2013 heat island effect measured in degrees Celsius according to the CIESIN data. A comparison between the BEST data and the reported CIESIN urban heat island differential suggests that the BEST data does not recover the extent to the heat islands derived in CIESIN. CIESIN results show that for some cities, such as Rio de Janeiro and Sao Paulo, the urban heat island effect is strong, with over 4°C difference between local urban and rural temperatures.

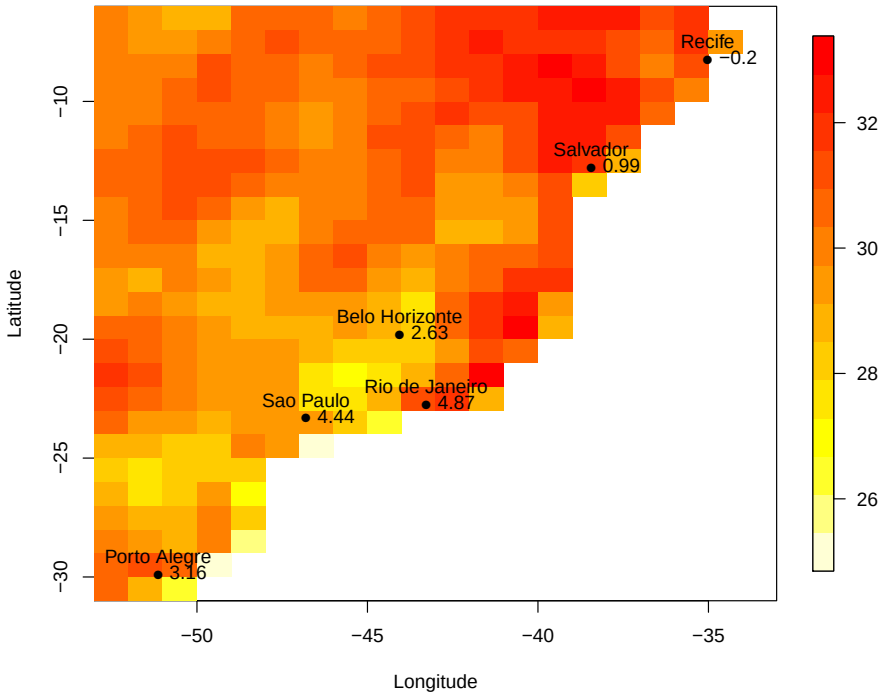


Figure 4.1: Mean summertime maximum temperatures in 2013 in Brazil according to the BEST dataset. The urban-rural differential calculated according to the CIESIN data is depicted in text.

In order to avoid potential bias in the BEST gridded dataset I took two approaches. First I conducted the analyses assuming that BEST accurately recovers the true urban heat effect. Second, I added the urban heat island differential pro-

vided in the CIESIN dataset to the BEST data assuming that the BEST dataset does not account for the heat island effect.

The costs of climate change in terms of reducing labor supply are weighed against the cost of implementing city-level cooling strategies. There are multiple reasons to implement such strategies including reducing heat-wave related mortality, reducing energy demand and reducing air pollution [35]. Reduction in labor supply represents only one these potential costs. Cooling strategies considered are green roofs, cool roofs and cool pavements. The costs of these policies are based on estimates in [37] adapted from the EPA [35]. The original estimates of the cooling effects of such mitigation strategies were based on models of specific American cities. I assume that these cooling effects are similar in the cities analysed here; however, further work should base the estimates of these costs on city–level modeling.

Appendix Table C.1 provides the cooling estimates and costs for each temperature reduction strategy. Four policy scenarios are considered: coof roofs, green roofs, cool pavement and cool roofs + cool pavement. The maximum temperature reduction is 0.8°C which is achieved through a combination of cool roof and cool pavement policies. The cost of these policies is calculated using estimates of the squared kilometer size of the urban area. Pavements are assumed to cover 35% of the city and rooftops are 25%, based on estimates in [36]. As in [36] it is assumed that policies are implemented in the present and costs of maintenance are not considered. This may be particularly problematic for green roofs which may require annual maintenance costs [35].

4.4 Results

4.4.1 The effect of temperature on labor supply in Brazil

Figure 4.2 shows the effect of temperature on labor supply using the preferred GCP specification. Three models are shown. The binned response and the fourth order

polynomial response were both fitted to data from Brazil. The global response curve was fitted to weighted data from the nine study regions. Interestingly, the polynomial Brazil response closely follows the global response for the portion where these temperatures overlap. The binned response estimates a sharper decline in labor than the polynomial response as temperatures exceed 35°C.

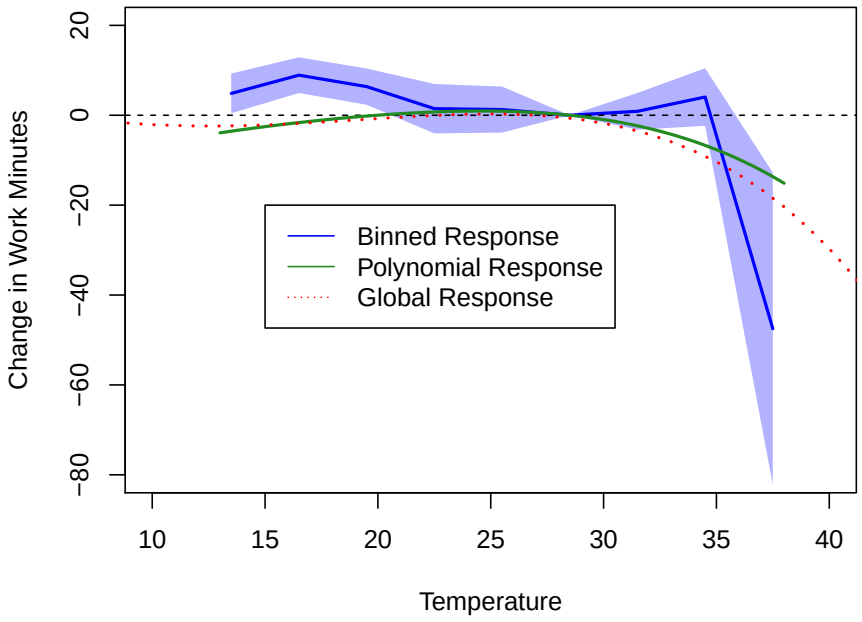


Figure 4.2: The effect of temperature on labor supply according to the GCP specification (using city by month-of-sample fixed effects). The results of a binned temperature regression (blue) and fourth order polynomial regression (green) are shown using the Brazil data. A fourth order polynomial response function for the global dataset is also shown (red dashed). The response functions for Brazil are plotted over the range of temperatures experienced in Brazil according to BEST data. The global response function is plotted from 10 to 40 degrees Celsius.

Sub-groupings of the population were analysed in Figure 4.3. The first plot shows high-risk versus low-risk. The second plot divides the data by sector and the third response looks at a few specific industries. High-risk groups include agriculture, mining, construction and manufacturing. The largest effect on labor supply is in the primary sector,where the hottest days decrease labor productivity by approximately

80 minutes relative to a 25°C day. The primary sector includes individuals working in agriculture and mining which experience the steepest declines according to the industry-specific results. These industries are typically considered "high-risk" as they often require outdoor work where an individual is exposed to the prevailing environmental conditions.

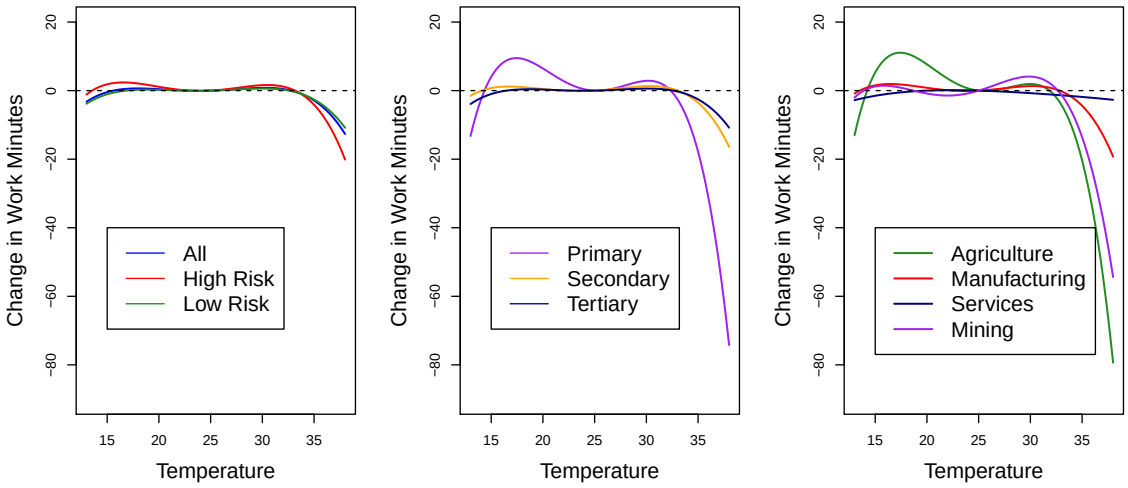


Figure 4.3: The effect of maximum daily temperature on working time for different employment groups and occupations fitted using a polynomial response function. Curves are normalized at 27°C.

The response of the high-risk group does not clearly differ from that of the low-risk group. This appears to be due to the inclusion of the manufacturing industry, which does not exhibit a strong response to temperature extremes and in numbers dwarfs the contribution of other industries. As all respondents are urban dwellers, the numbers working in agriculture (31,000) and mining (17,000) are relatively low compared to manufacturing jobs (730,000). The primary sector, in total, represents 0.8% of jobs in the sample. However, the tertiary and secondary sectors do exhibit a slight response – a reduction of approximately 20 minutes on the hottest days. The only sub-group that appears to be very minimally affected is the service sector.

The sample is divided such that responses can be assessed at the city level. Due to the large sample size, with close to 1 million observations per city, city level regressions

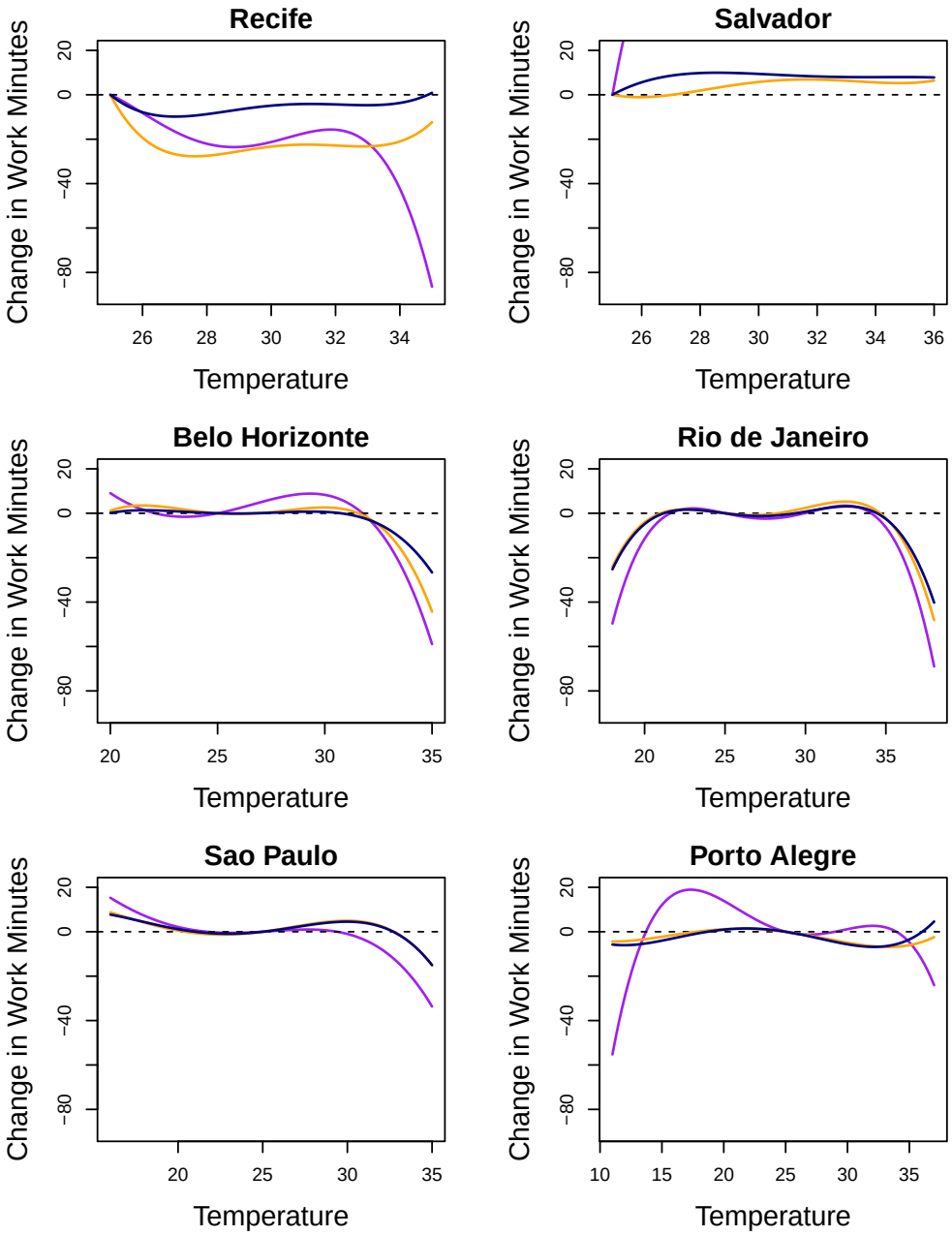


Figure 4.4: Separate fourth-order polynomial response functions fitted to each city by sector. Primary (purple), secondary (yellow) and tertiary (blue) sectors are shown. Temperature scales are changed to represent the temperatures experienced in each city. Labor scales are kept constant.

are sufficiently powered. For each city, primary, secondary and tertiary responses are assessed separately. Figure 4.4 shows the response functions for each city. For all cities except Salvador, the primary sector shows a decline in work minutes with

hotter temperatures. [4] Porto Alegre experiences cooler temperatures than the other cities and also sees a steep decline in primary sector labor during colder days.

Among other sectors there is heterogeneity in responses. In Rio de Janeiro and Belo Horizonte all sectors appear somewhat affected by temperature changes. There are multiple possible reasons for this difference; however, the heating effect of the urban heat island is one possible cause. If a 35°C day is actually experienced as a 39°C day due to the urban heat island, then even tertiary sector employees may be affected. Other cities, such as Recife, with no urban heat island effect, do not see primary and tertiary sectors affected at the exposed temperatures.

Robustness checks are shown in the appendix. Estimating the results based on temperature in the following week do not appear to show a strong pattern (Appendix Figure C.8) suggesting the model is not mis-specified. Lagged temperature also does not show a strong response (Appendix Figure C.9), though a slight uptick at higher temperatures may be indicative of temporal displacement in working hours, though the size of this change is much smaller than the loss in work experienced as a function of temperature.

4.4.2 Projected changes and costs of changes to labor supply driven by climate change

The main polynomial response function is used to estimate the potential change in labor supply due to climate change. Figure 4.5 shows this result for the six studied cities. Salvador experiences the greatest change in work time per individual whereas Porto Alegre experiences the least. These responses correspond to expected temperatures in these cities, i.e. Salvador will experience the hottest temperatures by the end of the century (Appendix Figure C.4).

[4] The response of the primary sector in Salvador appears to be anomalous and is perhaps driven by bias in the relatively low number of observations in this sub-group (7567).

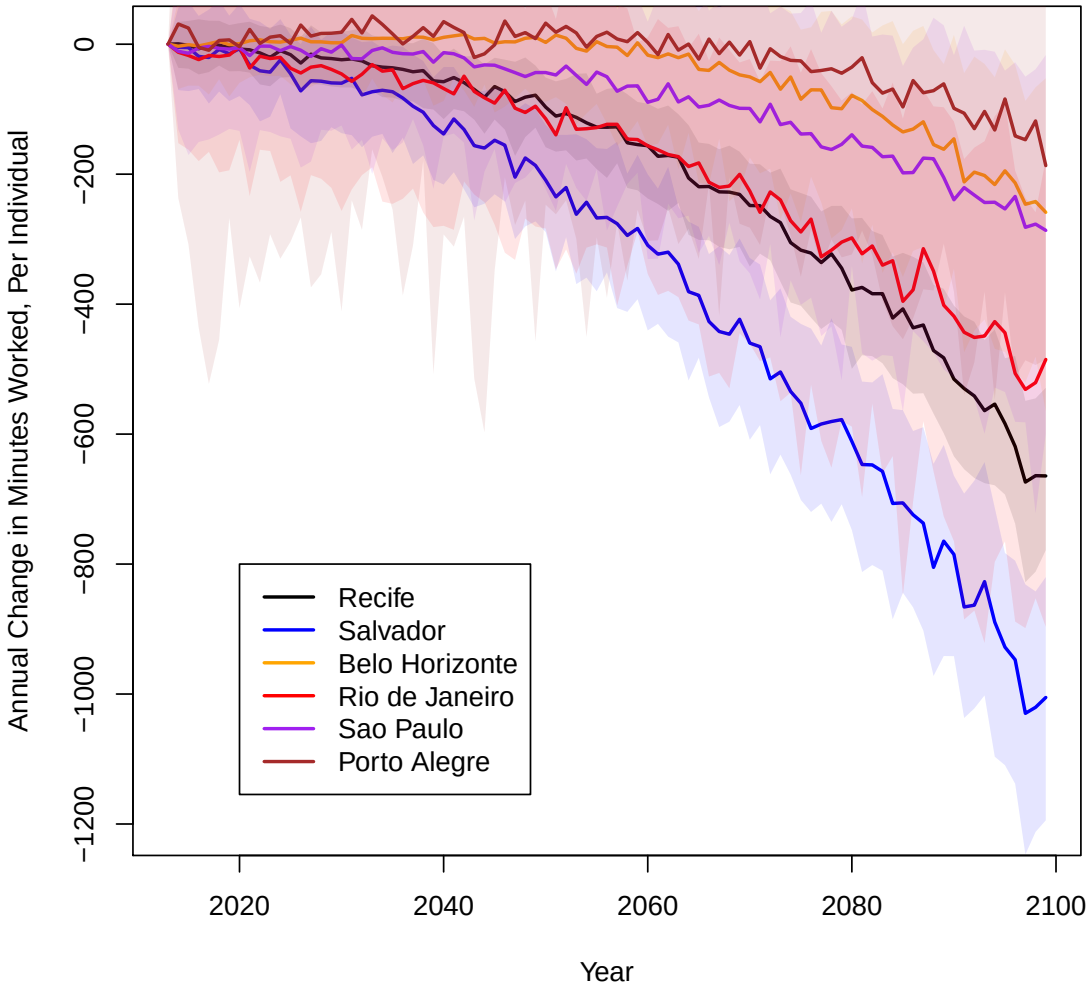


Figure 4.5: The projected effect of climate change on change in minutes worked for an individual living in each city. The results come from 100 simulation runs. Mean change in labor supply is given by the bold line and 95% confidence intervals are shown.

These projections do not account for adaptation or changes in the structure of the labor force. Analysis of the existing data suggests that the proportion of individuals within each of the three sectors (primary, secondary and tertiary) has not changed over the previous decade (Appendix Figure C.5); however, this may change in the next 100 years. Other adaptation measures, such as air-conditioning, improving energy supply

if blackouts occur and changes to modes of transportation that mitigate exposure are not considered.

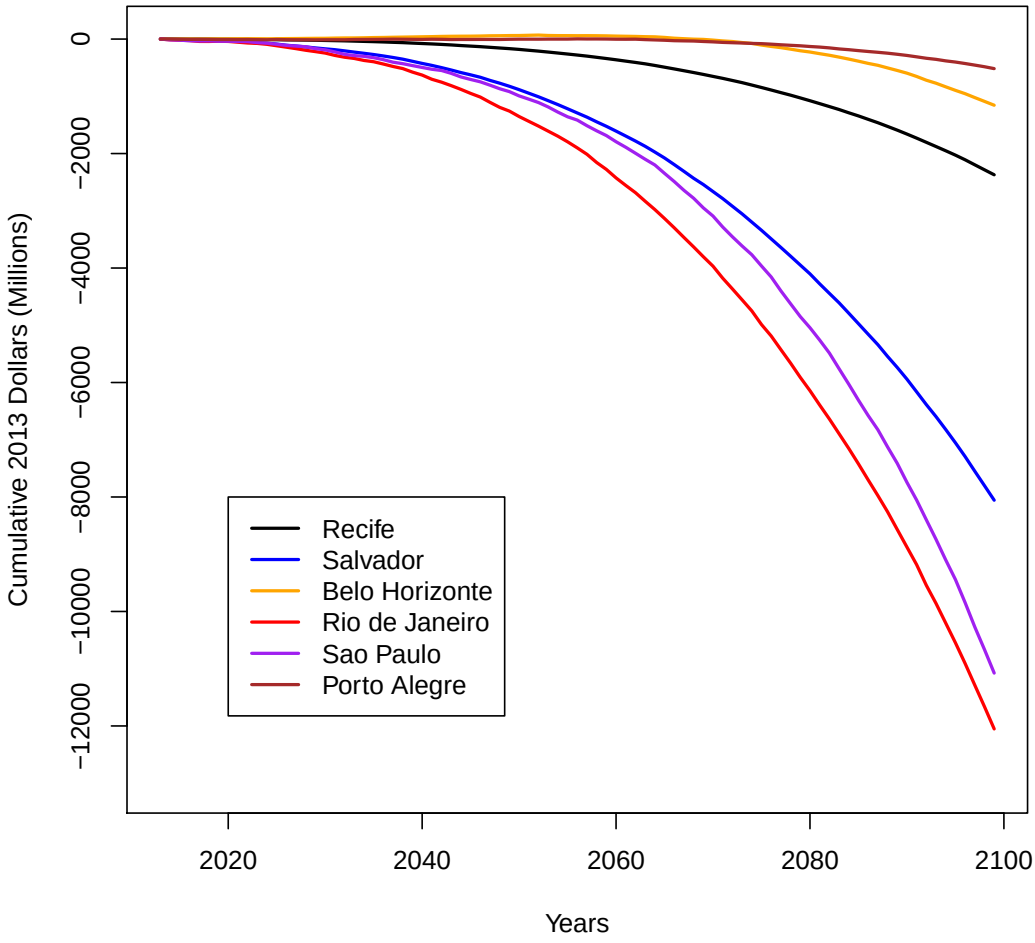


Figure 4.6: The projected economic effect of climate change due to changing labor supply across the six study cities, assessed using population projections for each city and valuing labor at the mean city wage.

The results suggest that labor supply may be reduced by up to 1000 minutes per individual by 2100 (approximately 17 hours annually). The economic consequences of such a reduction are assessed using the mean hourly wage. Lost labor is valued at the mean wage for each Brazilian city in 2013 and converted to dollars (Appendix Table C.3). The lost wages are multiplied by a projection of the number of individuals living

in each city. Urbanization continues to drive population growth in these regions over time, though this stabilizes towards the end of the century.

Figure 4.6 shows the cumulative cost in (millions of) 2013 US dollars of the lost labor supply for each city. Rio de Janeiro experiences the greatest losses (> $1.2 billion by 2100) : a function of both the large population size (7.7 million in 2100) and higher temperatures (Appendix Figure C.2). Sao Paulo also experiences large losses, though the city has almost double the population size (14.3 million in 2100) of Rio de Janeiro. Per capita losses are less extreme due to relatively cooler temperatures. Salvador experiences cumulative losses close to $8 billion by 2100. Although the city has a smaller population (3.48 million), it experiences the largest per capita declines in labor. Recife, Belo Horizonte and Porto Alegre see smaller changes, due to their relatively smaller populations. The total cumulative losses by 2100 are 35 billion (2013 US dollars).

4.4.3 Adjustments for urban heat islands

Urban heat islands may not be detected in standard gridded temperature datasets. An adjustment is necessary to incorporate the urban heat island effect into the BEST temperature dataset. Figure C.10 shows the new response function generated from the adjusted dataset. Temperatures in the adjusted dataset reach highs of 45°C and labor supply begins to decline as temperatures exceed 39°C.

The aim of the adjustment is to answer two questions. First, to what extent does the urban heat island exacerbate the negative consequences of climate change? The projection model allows for the hypothetical removal of the urban heat island effect from the temperature forcing such that a scenario with zero urban heat islands can be considered. This question is important for those planning the development of future cities, as newly designed cities may be able to minimize urban heat islands effects at the construction stage. The second question is to what extent can heat in

cities be mitigated by adaptation policy options and do these option make economic sense? In order to answer this question, four policy options were considered that were investigated by the EPA as potential heat island reduction strategies. These policy options could also be used to cool cities even in the absence of heat island effects.

Rio de Janeiro has the largest urban heat island effect out of the cities in the dataset. It also exhibits one of the strongest responses to temperature changes, with all three sectors clearly affected. Therefore policy changes to reduce urban heat islands will be of particular relevance to the city. Figure 4.7 shows the costs and benefits of reducing the urban heat island effect in Rio, evaluated solely in terms of changes to labor supply. The lines in the plot show the predicted cumulative damage from lost labor supply due to climate change up until the year 2100. The steepest (red) line is the current scenario based on the a +4.87°C urban heat island effect. All other lines show some reduction from this urban heat island effect. The colored lines show reduction is 0.5°C intervals. Interestingly, if the urban heat island effect is completely removed, there will be no effect of climate change on labor supply, i.e. the urban heat island in Rio de Janeiro pushes the city into the damaging range of temperatures. In fact, with a complete reduction in the urban heat island effect, climate change actually increases labor supply and the benefits thereof (blue line), due to decreasing cold days that also negatively impact labor (Figure C.10).

The cost benefit analysis for different policy options is shown in the inset bar chart. Three policy options appear viable: cool roofs, cool pavements and a combined approach of these two. Green roofs are much more expensive to implement and benefits from reduced labor supply damages alone do not appear to outweigh the costs in this case. Installing combined cool roofs and cool pavements could lead to a $6.5 billion reduction in losses and would costs an estimated $2.9 billion.

Another way to assess the policy choices is to calculate the date at which the policies, accrued benefits outweigh the cost of implementation. This may help with

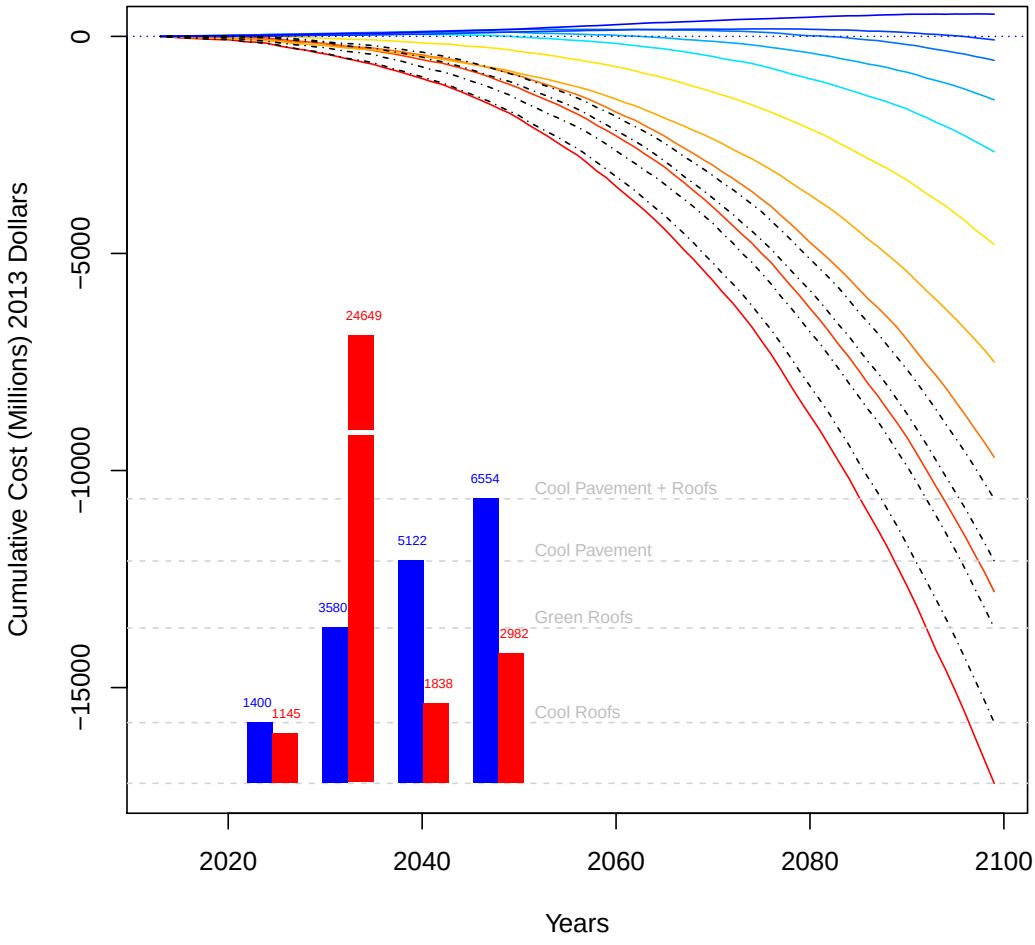


Figure 4.7: The costs and benefits of urban cooling in Rio de Janeiro. The colored lines show the projected cumulative damages of lost labor supply as the urban heat island effect is reduced from +4.87°C (red) in 0.5°C intervals to 0°C (blue). Dashed lines represent four policy scenarios to lower temperatures: cool roofs, green roofs, cool pavements and cool pavements + roofs. The inset bar chart compares the benefit of these policies (blue), calculated using the reduction in cumulative damages, to the costs of the policies (red), based on EPA estimates (see Methodology). The bars are labeled with the monetary value of these costs and benefits in US dollars.

longer term budgeting. Figure 4.8 shows these dates for the three viable policy options (green roofs are too expensive relative to labor damages to repay investment within the century). The shortest term repayment is achieved by cool pavements (56 years), then combined cool roofs and cool pavements (62 years) and finally, cool roofs alone

(80 years). All of these repayment time scales appear relatively longterm. However, it is important to remember that damages to the labor sector are only one of the multiple economic impacts of climate change. Mortality effects are one other obvious consideration.

4.5 Discussion

Temperature changes appear to be linked with labor supply in urban areas in Brazil. Particular industries, especially those associated with outdoor work such as agriculture, exhibit the strongest response. This supports the hypothesis that heat stress may cause agricultural laborers to change their labor supply as temperatures reach

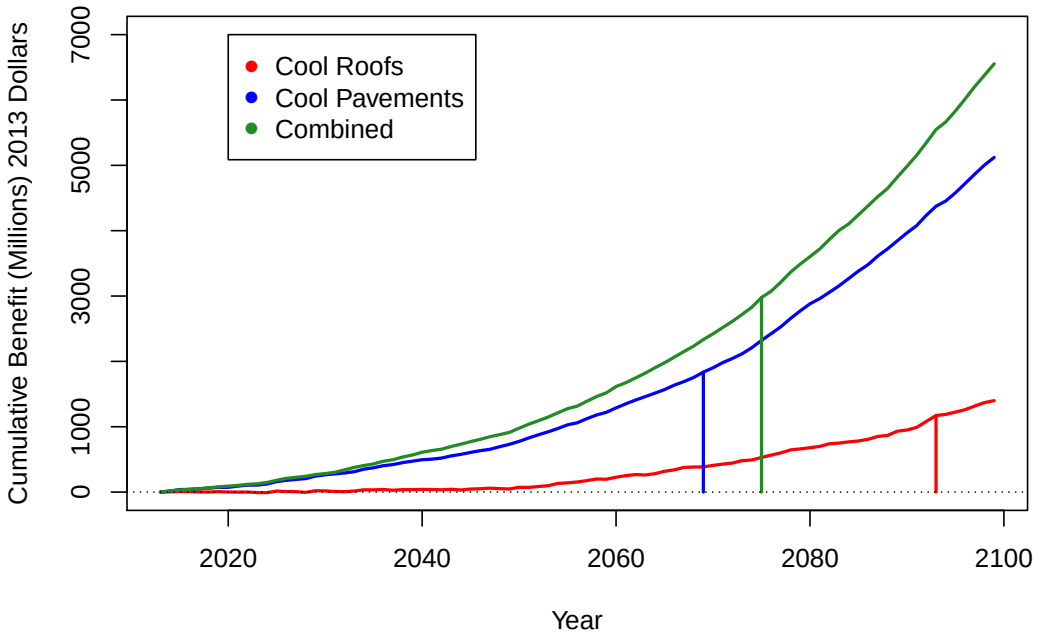


Figure 4.8: The cumulative benefits accrued from implementing each of three policy options. Benefits are in terms of recovered damages from otherwise lost labor supply. The date at which these cumulative benefit equal the costs of implementing the policy is shown with the vertical lines.

ranges that are detrimental to health. However, other causal mechanisms could potentially explain this relationship. For instance, agricultural losses during hot periods could change labor demand. One limitation of this study is that causal mechanisms cannot be further elucidated. Future work should build on the results found here to find the drivers behind the temperature-labor linkage.

As argued in Chapter 1, optimal models for prediction combine statistical and mechanistic approaches to consider future changes. The model developed in this chapter is primarily statistical and as such is limited in it's predictive capacity by concerns of external validity. In the long-run, multiple changes to the work force may alter the impact of temperature on labor supply. One obvious concern is that the sectoral make-up of the city work force may change over time. The typical narrative suggests that as nations advance economically, workers move from the primary sector to the secondary and tertiary sectors [83]. In the model for prediction developed here, sectoral make-up remains the same. This is based on two pieces of evidence. First, the percentage break-down between primary, secondary and tertiary sectors has remained unchanged over the previous ten years (Appendix Figure C.5). Second, the primary sector composes less than 1% of city workers. Such a level is argued to already be at a minimum and is unlikely to diminish further. Changes between the secondary and tertiary sectors are unlikely to alter the predicted responses as both sectors exhibit a similar climatic response.

Another consideration is that air-conditioning may limit negative responses in occupations where individuals work indoors. As cities warm and as incomes increase, air-conditioning may become a worthwhile investment in both the work place and at home. Longitudinal data on air-conditioning penetration is hard to find, however cross-sectional data is available (Appendix Table C.3). There is a large amount of regional heterogeneity in air-conditioning usage rates in Brazil. In Rio de Janeiro, approximately 24% of households had air-conditioning in 2010, however, in Belo Hor-

izonte just 1.6% of households have air-conditioning. It is hard to discern a cross-sectional driver for differences in air-conditioning uptake; there is no clear association with mean wage or mean temperature. Nevertheless, the approach developed by the Global Climate Prospectus to include mean temperature and mean income as interaction terms appears to be the best way to capture adaptation responses that scale with these two variables; air-conditioning may be one such response.

Infrastructure changes may be one causal mechanism linking temperature shocks with the observed changes in labor supply. The failure of transport systems, or electricity systems could cause whole businesses to shut down on hot days. Blackouts may be behind sector-wide changes to labor supply. Future studies should attempt to either control for this, or model the likelihood of such responses. Analysis of different income levels suggests that individuals who receive a lower monthly wage are most like to reduce their labor supply in response to temperature shocks (Appendix Figure C.11). This relationship holds even when the primary-sector (e.g. agriculture) is excluded. This suggests that lower income workers are more likely to be in jobs with environmental exposure, where hot days increase the cost of working.

Cities can implement cooling policies such as cool roofs or cool pavements to partially mitigate the negative consequences of climate change. These policies appear expensive: conditional on some simplifying assumptions, I estimate that even the cheapest cooling policy in Rio de Janeiro would cost 1.14 billion US dollars. This represents approximately 1% of Rio de Janeiro's GDP. [5] Without discounting future losses, the costs of implementing these measures may outweight the benefits. However, including a 3% discount rate makes all policy options too expensive when assessed against labor damages alone (Appendix Figure C.6). Damages to the labor sector should be considered alongside the other negative consequences of climate change in assessing the benefits of reducing urban heating.

[5]Using 2013 GDP data accessed from Instituto Brasileiro de Geografia e Estatstica (IBGE).

The creation of urban heat islands is the result of the structure and the materials used when constructing it. Cooling policies can help reduce urban heat island effects in existing cities, but new urban developments may also be able to take into account the impact of materials used at the construction stage. This may be a more cost-effective option for mitigating the negative effect of urban heat islands where possible. In locations where rapid urbanization is taking place, city planners should attempt to reduce future urban heating. The results for Rio de Janeiro suggest that a complete reduction in the urban heat island could turn the effect of climate change from negative to positive, at least for the labor sector. However, specific modeling is required to assess this result for each new city.

Chapter 5

Conclusion

Climate-impacts, as a new field of study, is growing rapidly. The increasing availability of high resolution data on both socioeconomic indicators and climatic variables is generating a wave of research, with new discoveries of previously unknown linkages frequently uncovered. Whilst data-driven analyses are common, methodological advances are also occurring, particularly as conversations take place across disciplines and tools from different fields are merged.

This thesis considers the implications of climate change for three separate outcomes. Although these projects are distinct, methodological innovations regarding the prediction of climate change impacts remain a common thread. Such predictions necessarily come with limitations, and of course, room for improvement.

Uncertainty

A thorough treatment of uncertainty is not provided in any of the studies presented in this thesis, and remains an avenue for further research. Treating uncertainty appropriately within climate-impacts work is complex. Predictions carry uncertainty not only in the statistically estimated *parameters*, used to characterize a particular model, but also in terms of how the input *variables* within the model will change

over time. This distinction between parameter uncertainty and variable uncertainty is important. I define a variable as an input into a model that is usually measured directly (even if it is only possible to measure at one point in time) and a parameter as an "embodiment of a concept" that "does not have the reality of a directly recordable variable" [19]. Parameters within a statistical model are estimated relationships between variables. Variables within a climate-impacts model may include temperature, population size, the number of susceptible individuals, for example. Parameters may represent a specific concept, such as the transmission rate, or a more abstract concept, such as the slope of a line relating temperature changes to hours worked.

Parameter uncertainty

The statistical uncertainty typically considered when fitting a model, is based on uncertainty in the estimated parameters. It is defined as $var(\hat{Y}/model)$ which represents the variance in the predicted outcome $\hat{Y}$ given the model used to generate the prediction. The method for calculating this will depend on the model used, but could involve sampling from parameter distributions using Monte Carlo approaches, for example. Exploring the effect of this sort of uncertainty is particular enlightening for more complex systems, where results do not scale linearly with parameter changes. Note that this uncertainty is fixed at the point at which the model is fitted to data, and is not time-dependent.

Variable uncertainty: projected social changes

The value that social variables and climate variables will take in the future is uncertain, yet these are core components of prediction models (see Introduction Figure 1.2). For example, population growth rates could change dramatically towards the end of the century, depending on location-specific factors. In Chapter 2, I develop a range of population growth scenarios when predicting HIV prevalence. I find that variations

in population growth rates have a strong effect on the predicted number affected by HIV, as the climate changes (see Appendix Figure A.7).The hybrid statistical-mechanistic model used in this chapter provide a simple framework for testing the effect of alternate growth scenarios.

Another advantage of leveraging a hybrid approach is that it may highlight which social variables are important for future predictions. Because a mechanistic model more explicitly depicts the system governing the relationships between variables, the model can be used to guide uncertainty analysis. In some cases additional variables may become important over time, even if they were not included in the model in the first place. For example, migration numbers could affect labor supply in the future, even if they do not now. Typically, model selection is based on *parsimony*, but a distinction should be made between *present parsimony* and *future parsimony*: the best model for predicting current changes, even in the hybrid statistical-mechanistic framework, may not be the best model for considering the future.

Variable uncertainty: projected climatic changes

The extent to which the climate will change in the future remains deeply uncertain. In Chapter 2, I look at the results of HIV changes based on a range of representative concentration pathways (RCPs). These are scenarios of future greenhouse gas concentration trajectories which are named based on radiative forcing changes in 2100. RCPs are converted to temperature (and other climate) changes by general circulation models (GCMs).

The uncertainty within, and between, GCMs is well documented [135, 72]. Tebaldi documents four sources of uncertainty: uncertainty in initial conditions, uncertainty in boundary conditions, parameter uncertainty and structural uncertainties [135]. Uncertainties due to initial conditions are driven by the chaotic nature of the weather system. Uncertainties in boundary conditions covers the application of different green-

house gas scenarios based on differing population growth and development predictions. Parameter uncertainty, in the climate model case, is due to simplifications of small-scale processes that cannot be resolved otherwise in the model and summarized by a parameter. Structural uncertainty captures core decisions in which processes to model and which to ignore, given the complexity of the climate system.

In sum, the resulting uncertainty in future climatic changes, especially when downscaled to particular regional changes, may be large. Exploring what the range of climate projections may mean for climate-impacts is important from a risk analysis perspective. In particular, although extreme draws from a climate distribution may be unlikely, the outcomes of such draws could be damaging enough that measures should be put in place to protect against possible negative outcomes. The framework developed in Chapter 3 to investigate the effect of dryer conditions on the transmission of airborne, immunizing, infectious disease seems one possible avenue to explore the effect of extremes. The simulations run in Chapter 3 use mean changes in humidity, however, periods of extreme drying or "dry waves" might have a particularly large effect on transmission that could result in a significant increase in cases within a short time period. Public health officials planning vaccinations or other adaptations may need to assess the risk of such extremes. Using the core model developed in Chapter 3, simulations could be run for a host of similar infections and potentially others, to consider such changes.

Dynamic Models

Chapters 2 and 3 both incorporate considerations of the nonlinear dynamics of disease in long-term predictions. Considering dynamics has, to some extent, been left out of the climate-impacts literature. This may be to the detriment of the field. Evidence suggests that if the causal relationships within a complex system are appropriately tested for, linkages may be found that would have been undetected by traditional

approaches [130]. Hybrid statistical-mechanistic models are particularly well suited for analyzing nonlinear complex dynamic systems. I suggest two areas of the climate-impacts literature that may benefit from such an approach: mortality and migration.

The effect of temperature on mortality is well-established in multiple disciplines [30, 45]. However dynamic responses are both important and understudied. With global warming expected to bring more frequent and longer-lasting heat waves [86], the effect of these periods on mortality should be analysed. The framework of disease models wherein a proportion of the population is viewed as *susceptible* may have relevance to mortality studies where certain subsets of the population are at risk. The dynamics of how this population changes over time and how this might affect mortality during a given heat wave period are important to understand. For example, the first hot day within a heat wave may kill a certain number of vulnerable individuals, however, the second hot day may not have the same total effect on mortality, even if the same temperature is experienced (because the number of vulnerable individuals has been reduced). Over longer periods of time, increasing numbers of the population will become vulnerable, due to aging or sickness, such that a heat wave experienced in the subsequent year may again have a high effect on mortality. An alternative hypothesis is that cumulative days of warming may increase the vulnerable population by exacerbating underlying illnesses, for example. Daily heat-related mortality rates in this case may rise over cumulative hot days. The crucial concept is that mortality rates due to a particular temperature are not just a function of location, but also a function of the history of prior warming, both in terms of number, and temporal structure, of prior heatwave days. Understanding these dynamic responses may improve models of future heat-related mortality, and also allow for appropriate response planning.

The effect of climate shocks on migration is also relatively well established [82, 12, 91, 52, 51], though data limitations have hindered advances in this field. Theories

of migration also suggest that there are dynamics at play [55, 92, 84]: as the flow of individuals to a particular location increases, wages at this receiving location decrease, and reduce the benefits of future moves, conversely, the cost of moving is diminished by having family ties (earlier migrants) at this location. Little work has attempted to formally integrate climate shocks into such a model, statistical techniques could be used to fit these models to time series data on migration numbers.

Adaptation policy

Climate change will have wide-ranging effects on human society, as suggested by the increasing number of linkages found in the climate-impacts field. The implementation of adaptation measures may be able to mitigate some of the negative consequences of climate change. However, policy makers have finite resources with which to invest in adaptation measures, even if long-run costs outweigh benefits. Future work should begin to look across the output of the climate-impacts field, to identify an order of priorities for potential interventions. There are multiple dimensions that need to be considered in this case: the severity of the outcome, the probability of the outcome, the timescale of the outcome (both start point and duration) and the cost of intervention. Such an exercise may occur at multiple-levels: from city-level adaptation plans, to national or multi-national level decision-making.

There appear to be synergies between exercises to predict and value climate change impacts and the exercise required to assess potential adaptation options. Climate-impacts studies can answer questions regarding the severity, timing and cost of an outcome. The prediction models used in these studies can also be used to explore adaptation measures. In Chapter 2 I use the HIV model to assess one potential intervention: increasing antiretroviral coverage. Embedding this analysis within the framework of an existing prediction model is a relatively straightforward approach. However, there are many other ways in which climate change will affect Sub-Saharan

Africa. It is difficult to assess what adaptation measures should be a priority, given finite budgets. For instance, climate-driven changes to malaria prevalence, or threats to health from heat-waves, may be much more of an immediate concern than HIV for some locales. More works needs to be done to look across these potential impacts to help develop targeted policy.

A strong framework exists for thinking about mitigation policy: marginal abatement curves, potential co-benefits for reducing emissions and stabilization wedges, for example [103]. The framework for adaptation is more loosely defined. Clearer thinking on this front might motivate research efforts in this direction. Some of the mitigation framework could be borrowed for this use: the co-benefits of adaptation policy across sectors could be considered, as could a marginal adaptation cost-curve (though adaptation benefits may be harder to quantify).

Data

The rapid rise in data availability, coupled with advances in technology to handle large datasets, is generating a wave of new techniques for data analysis. These new techniques, particularly those of "machine learning", fall under the bracket of *statistical models* discussed in the Introduction. I argue that statistical models are good for in-sample and near-term prediction but insufficient for long-run predictions. Some of the machine learning models lie even further along the "statistical" spectrum than the linear regression methods discussed earlier: these models can be indiscriminately fitted to large datasets of many features (variables) without scientific/causal interpretation. Despite this, there are several opportunities for leveraging these techniques and approaches for climate change impacts prediction, in the short-term.

Machine learning methods have been applied to combine large satellite datasets of atmospheric variables and land surface variables to generate predictions of malaria prevalence spatially, based on observations in particular locations [47]. A similar ap-

proach could be used to provide short-term predictions (i.e over the coming months) of malaria, or otherwise, that could have major implications for individuals in endemic regions. In addition, this approach could be applied to other infectious diseases, and potentially other climate impacts [57]. Research shows that in some industries the availability of short-term climate forecasts, at least for ENSO related perturbations, allowed firms to completely adapt to these shocks [126]. Similarly, short-term predictions of potential changes to disease risk, may be valuable information for public health officials and individuals who can take actions to minimize their risk of exposure. Such results could not be used to estimate a social cost of carbon, but could be developed as an adaptation tool as climatic changes alter local risks.

Summary

The work presented in this thesis attempts to make contributions to the climate-impacts literature by both exploring new relationships and developing new methods. I find that across multiple arenas, the fate of human beings remains ultimately tied to the environment. The extent to which the predictions generated here will eventually come to pass, depends on whether we begin to acknowledge the scope of our socio-environmental interdependence. Action is required, as soon as possible, to avoid destructive changes.

Appendix A

Climate change drives modeled HIV prevalence

A.1 Modeling HIV

There are multiple approaches to modeling HIV, reflecting the complexity of the social, behavioral, cultural and biological processes that govern the dynamics of this disease. These models range from simple compartmental model's where incidence is governed by sums of flows between states [1], to complex agent-based models where each individual is traced across time in a network of interactions [2]. Choosing an appropriate model is a tradeoff between parsimony given data constraints and the model's ability to generate meaningful predictions.

In this chapter, the objective of the model is to capture the potential effect of climate change on HIV prevalence over multiple decades across multiple countries. Due to the scope of this study, a model is required wherein the core structure is

[1] e.g. Granich, Reuben M., et al. "Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model." The Lancet 373.9657 (2009): 48-57.

[2] e.g. Jenness, S., et al. "EpiModel: Mathematical modeling of infectious disease (2017)." URL https://cran. r-project. org/package= EpiModel. R package version 1.8: 155.

simplistic enough to reflect a potential commonality in the dynamics of HIV across regions. Furthermore, to capture the long-term dynamics of HIV, the model must fit country-level prevalence data and generate incidence numbers to assess total cases.

The chosen model is a compartmental model where individuals are defined as high-risk or low-risk. This allows the incorporation of temperature as influencing the proportion of individuals in the high-risk group, which is supported by the statistical analysis on risk factors. The temperature dependence is parameterized based on estimated statistical results. The same model is fit to prevalence data for every country. The model fit is governed by varying five parameters: the partne- change rate of the high-risk group, the partner-change rate of the low-risk group, the proportion of individuals in the high-risk group, the rate of supply of new susceptibles and the dampening effect of antiretroviral usage on transmission. Limiting the fit to changes in these five parameters allows the model to capture variations between countries while avoiding further unnecessary structural assumptions.

It should be noted, though, that the model is simplistic and it is hoped that future work in this field will attempt to incorporate temperature dependence in more complex, existing models of HIV spread. There are several clear extensions. First, the model does not take into account concurrency, which is often cited as one of the primary drivers of the rapidity of HIV spread in Sub-Saharan Africa,[3] though recent work calls into question this theory.[4] Second, the model does not take into account the age structure of infections, despite the fact that young adults account for most new infections. Dividing the population into high- and low-risk groups may partially account for this. The model also does not account for other factors that may influence HIV such as co-infections that may increase susceptibility, male circumcision or perinatal HIV transmission.

[3] Kretzschmar, Mirjam, and Martina Morris. "Measures of concurrency in networks and the spread of infectious disease."Mathematical biosciences133.2 (1996): 165-195.

[4]Lurie, Mark N., and Samantha Rosenthal. "Concurrent partnerships as a driver of the HIV epidemic in sub-Saharan Africa? The evidence is limited."AIDS and Behavior14.1 (2010): 17-24.

A.1.1 Bayesian MCMC

The model given in Section 1.3 is fit using a Monte-Carlo Markov-Chain with a Metropolis-Hastings algorithm. This process is a method for sampling from the prior distribution of parameters such that points in parameter space that are more probable are sampled more frequently. All parameters have uniform prior distributions. The high-risk and low-risk groups are defined by their relatively higher and lower partner change rates. The partner-change rate of the high-risk group is bounded [1,15] and the partner-change rate of the low-risk group is bounded [0,4] and the partner change rate of the high-risk group always exceeds the partner-change rate of the low-risk group. The proportion of individuals in the high-risk group is bounded [0,1], as is the efficacy of antiretrovirals on reducing transmission. Finally, the supply of new susceptibles is bounded [0.01,0.07]. This is the rate at which individuals in the population become sexually active and therefore at risk of contracting HIV. The range of values reflects possible population growth rates. Note that this does not explicitly equal the population growth rate, as the number of sexually active individuals does not necessarily grow at the same rate as the total population.

The model was run from 1960 assuming that HIV emerged around this time.[5] The model was initialized such that the total population size equals the population in the country in 1960 and there are 100 individuals infected with HIV in the high-risk group. There are several fixed parameters in the model. The per-partnership transmission rate is set at 0.05. This is based on the parameter used in Vynnycky and White, which is based on data from Rakai, Uganda [6]. Using a per-partner transmission rate as opposed to a per-act transmission rate increases bias in the model; however, it vastly simplifies the modeling process. The mortality rate of the population at sexual

[5] Zhu, Tuofu, et al. "An African HIV-1 sequence from 1959 and implications for the origin of the epidemic."Nature391.6667 (1998): 594.

[6]Wawer, Maria J., et al. "Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda."The Journal of Infectious Diseases191.9 (2005): 1403-1409

debut is set at 1/35, the mortality rate of AIDS is 1/1 and the rate of progression of HIV to AIDS is 1/9 representing the on-average ten-year time from infection to death [7].

The prevalence at each point in time was calculated from the model (i.e. the sum of those across all infected states divided by the total number in the model). A Gaussian likelihood function was used to compare the modeled prevalence to observed prevalence data. I make an adjustment to the likelihood function to downweight the contribution of earlier observations. This accounts for potential under-reporting of HIV in earlier time periods.

The Metropolis-Hastings algorithm samples from the prior distribution at each iteration with the sampled value being only dependent on the current values. The new sampled values are accepted with a probability proportional to the ratio of the posterior of this new sample to the posterior of the current sample. The proposal standard deviation (which governs the range of the distance between draws within the algorithm) was optimized such that the acceptance rate was kept between 20-50% across countries.

The model outputs the total number of HIV positive individuals in the population for every year. The incidence, i.e the number of new cases, was derived from this model output. The sum of the incidence for every year up until 2050 was used to calculate the additional cases due to climate change. It was also used to calculate the potential reduction in these cases due to an increase in antiretroviral coverage.

A.2 Tables and Figures

[7] Todd, Jim, et al. "Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy." (2007): S55-S63.

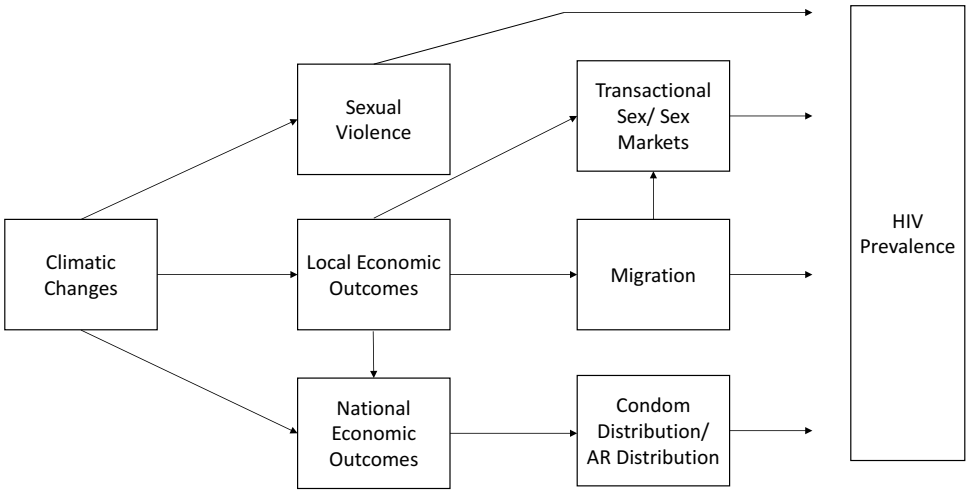


Figure A.1: The potential pathways linking temperature changes to HIV prevalence changes.

VARIABLES	(1) HIV+	(2) HIV+	(3) HIV+	(4) HIV+
6 Year Temperature	-0.000767	0.00414**	0.00419**	0.00419**
	(0.000832)	(0.00156)	(0.00181)	(0.00184)
Observations	415,337	415,334	415,334	415,334
R-squared	0.066	0.096	0.096	0.096
Country FE	X			
Year FE	X	X		
Gridcell FE		X	X	X
County*Year FE			X	X
Clustering	Country*Year	Country*Year	Country*Year	Country+Year

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.1: Alternative specifications for the main regression result. Model 3 is reported in the main text.

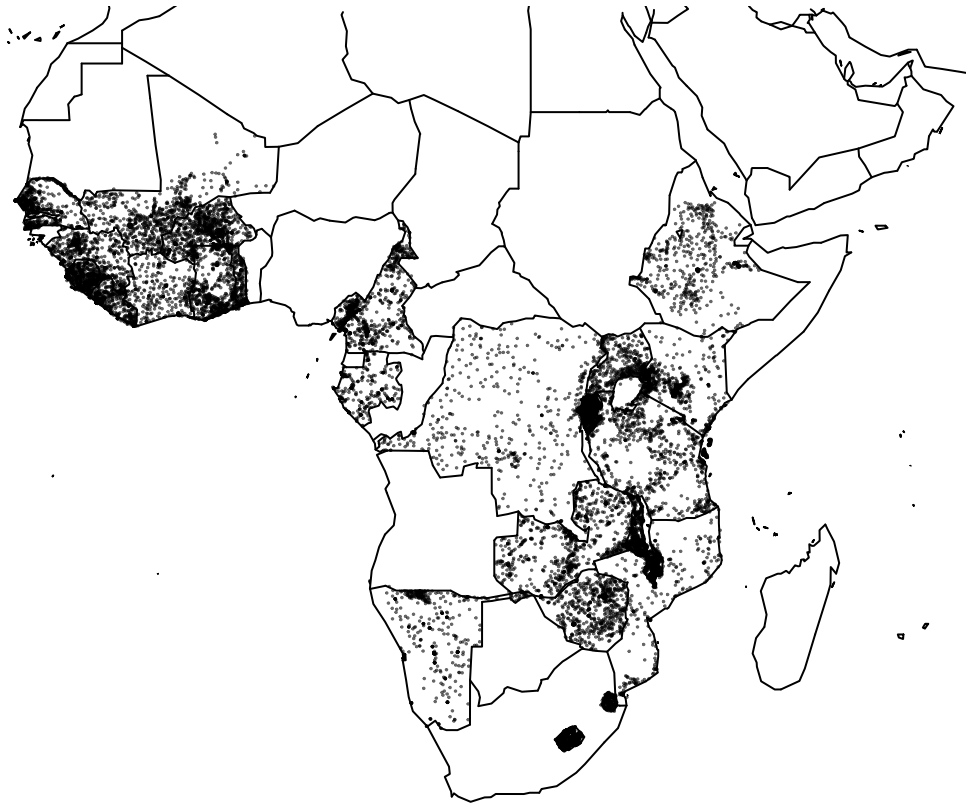


Figure A.2: Data: the locations of Demographic and Health Survey sites.

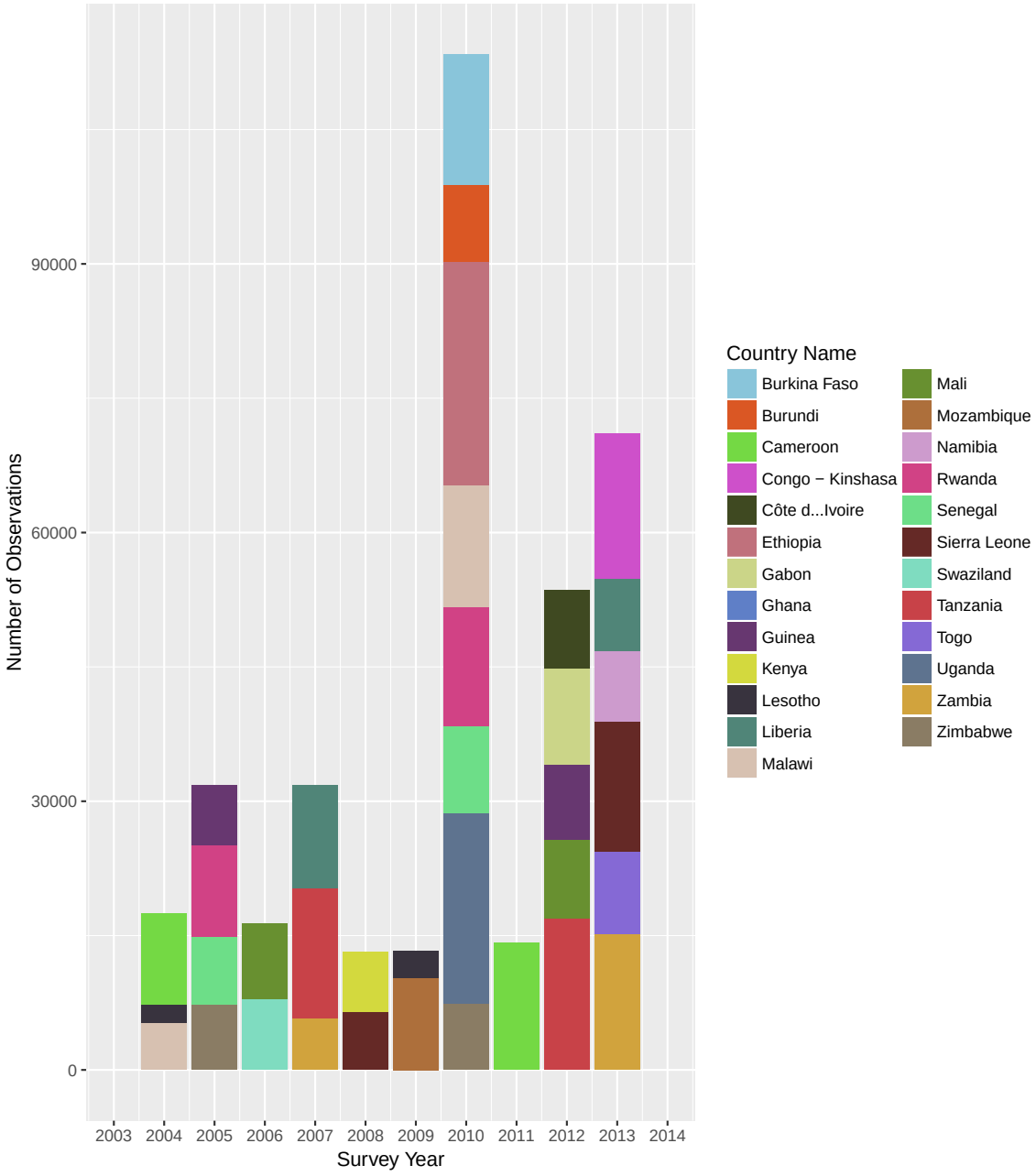


Figure A.3: The number of observations, country and timing of each Demographic and Health Survey used in the sample. Note, not every country with DHS data is used in the final climate change projections due to other data limitations.

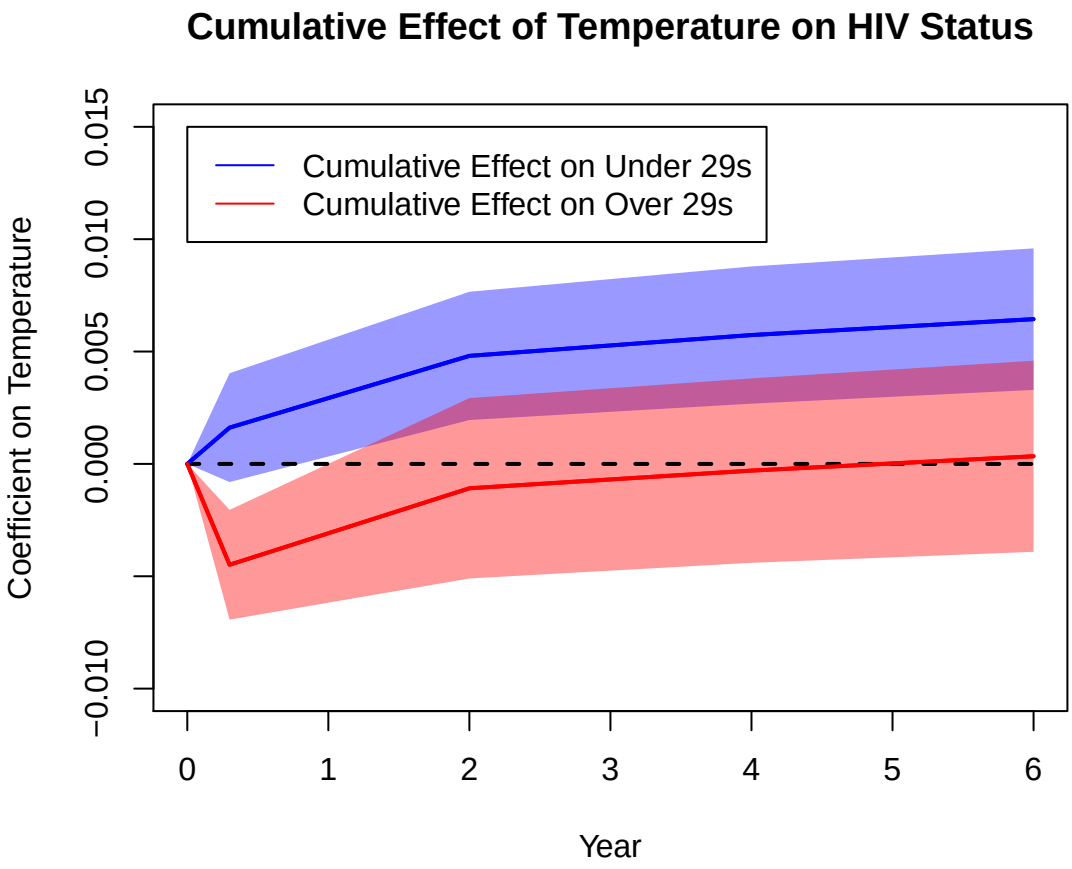


Figure A.4: The effect of temperature changes on HIV prevalence when temperature is averaged over different time periods. Regressions were run for three months, two years, four years and six years. The blue line is for adults under the age of 29 (the mean age in the sample) and the red line is for adults over 29. Over-29s see a brief decline in HIV prevalence with short-run temperature increases. This could be the effect of increased mortality for HIV/AIDS sufferers during hot spells. The younger age group shows a positive effect of temperature on HIV that is larger when looking at changes in temperature over longer periods of time.

VARIABLES	(1) HIV+	(2) HIV+
3 Month Lead Temperature	0.000129	
	(0.000656)	
3 Month Lagged Temperature	-0.00132**	
	(0.000625)	
3 Month- 6 Year Lagged Temperature	0.00510**	
	(0.00241)	
6 Year Lead Temperature		-0.000555
		(0.00421)
6 Year Lagged Temperature		0.00588
		(0.00586)
Observations	399,103	139,311
R-squared	0.097	0.131
Gridcell FE	X	X
Country*Year FE	X	X
HIV Weights	X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.2: The main econometric model was test for misspecification by regressing future temperature on present HIV prevalence. This was tested two ways. First, I tested the effect of temperature three months in the future relative to lagged temperature over three months and six years (Model 1). Future temperature is not significantly associated with increasing HIV prevalence. Second, I tested, the effect of six-year temperature in the future on current HIV prevalence (Model 2). This significantly reduces the sample size as the second round of most surveys is conducted within the last six years such that the sample is reduced to just surveys where both rounds are completed prior to 2010. The magnitude of the coefficients suggests that future temperature is not related to present prevalence, though neither coefficient is significant. Thus the main result appears robust.

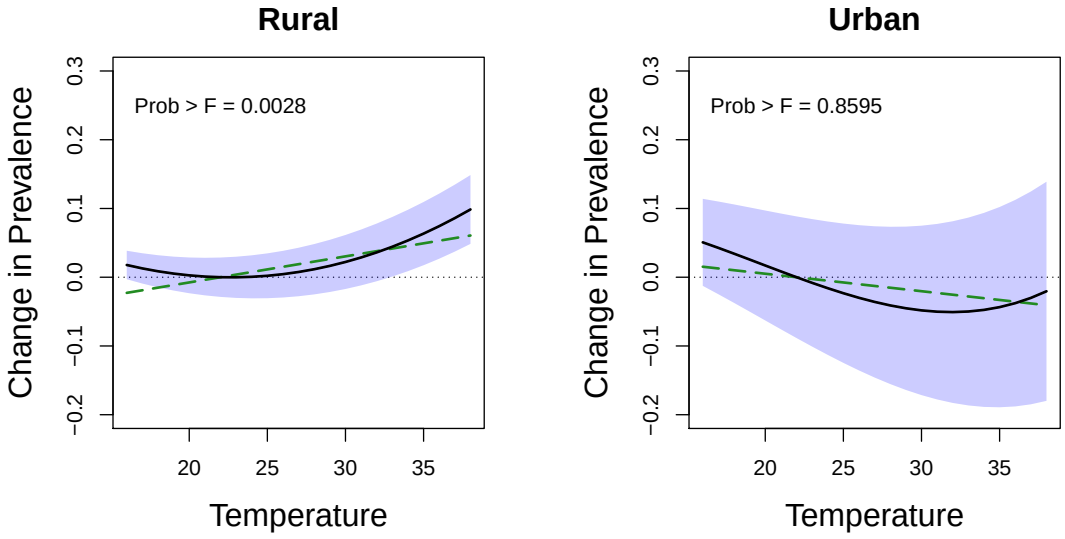


Figure A.5: The effect of temperature (averaged over a six year period) on HIV prevalence for rural and urban locations separately. The results of two models are shown: a third order polynomial response function, with confidence intervals in blue, and a simple linear response (dashed green line). Joint significance tests for the polynomial models are given in the top left hand corner. The response is significant in rural areas but not in urban areas.

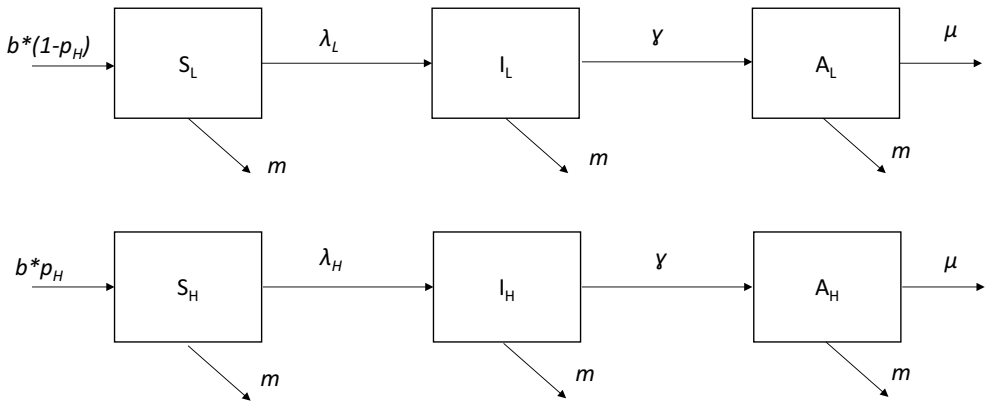


Figure A.6: A diagram representing the HIV model. Boxes represent state (stock) variables and arrows represent flows between states.

Country Name	Prevalence Change
Burundi	0.34%
Burkina Faso	0.22%
Cote d'Ivoire	1.82%
Cameroon	1.94%
Gabon	0.3%
Ghana	0.93%
Guinea	1.09%
Kenya	0.79%
Liberia	0.65%
Lesotho	3.89%
Mali	1.13%
Mozambique	3.48%
Malawi	0.64%
Namibia	3.75%
Rwanda	1.48%
Senegal	0.27%
Sierra Leone	0.92%
Swaziland	4.39%
Togo	0.91%
Tanzania	0.99%
Uganda	2.12%
Zambia	3.74%
Zimbabwe	2.03%

Table A.3: The model results under the RCP8.5 warming scenario. Column 2 gives the difference in prevalence, in terms of percentage points, between the RCP8.5 scenario and the no-warming scenario in 2050.]

Country.Name	c_H	c_L	p_H	b	k
Burundi	7.29	0	0.041	0.01	0.929
Burkina Faso	8.69	0.001	0.035	0.01	0.943
Cote d'Ivoire	8	0.007	0.108	0.011	0.491
Cameroon	8.6	0.064	0.098	0.025	0.066
Gabon	6.32	0.007	0.137	0.013	0.939
Ghana	8.06	0.017	0.039	0.016	0.388
Guinea	7.74	0.012	0.037	0.044	0.035
Kenya	9.52	0.002	0.224	0.01	0.872
Liberia	6.51	0.001	0.033	0.01	0.874
Lesotho	8.91	0.552	0.389	0.038	0.281
Mali	9.07	0.008	0.031	0.042	0.078
Mozambique	9.01	0.298	0.176	0.015	0.408
Malawi	8.87	0.009	0.348	0.011	0.953
Namibia	7.29	0.03	0.399	0.013	0.463
Rwanda	8.77	0.002	0.106	0.011	0.174
Senegal	6.1	0.001	0.003	0.016	0.046
Sierra Leone	5.87	0.001	0.03	0.01	0.959
Swaziland	9.36	1.046	0.384	0.038	0.117
Togo	7.1	0.001	0.094	0.01	0.961
Tanzania	9.32	0.014	0.154	0.02	0.928
Uganda	11.25	0.005	0.231	0.011	0.406
Zambia	9.88	0.184	0.283	0.015	0.035
Zimbabwe	9.55	0.008	0.494	0.011	0.795

Table A.4: The output of the model-fitting procedure for each country. Five parameters were fit using a Bayesian Monte-Carlo Markov Chain procedure using a Metropolis-Hastings algorithm with 25000 iterations.

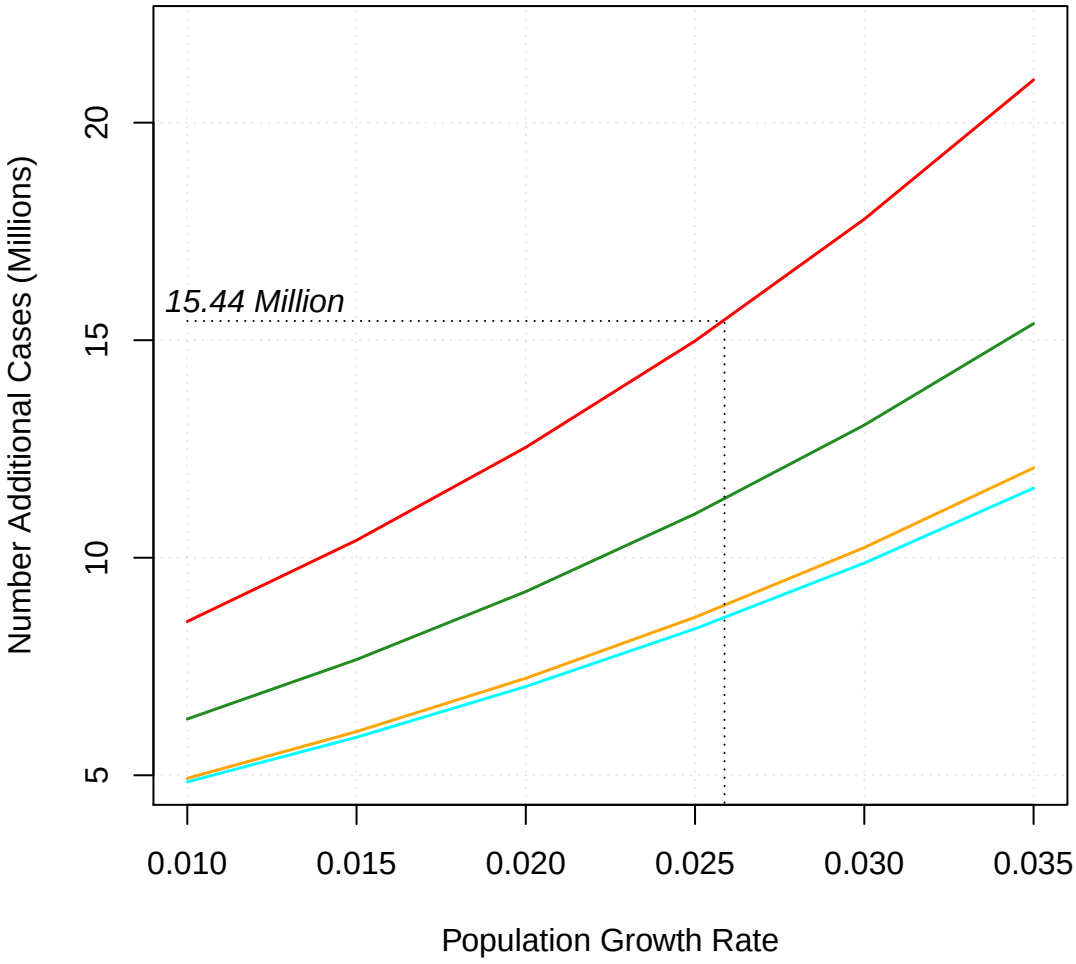


Figure A.7: Modeled incidence between the present day and 2050 over a range of population growth rate scenarios. The red line represents warming under an RCP8.5 scenario, with cyan, green and orange representing RCP2.6, RCP4.5 and RCP6.0 respectively. Note that the RCP6.0 scenario predicts lower warming when compared to the RCP4.5 scenario prior to 2050 but increases more rapidly towards the end of the century. The population growth rates range from 1% to 3.5%, representing the range of current (2015) country level growth rates according to World Bank data. The dashed line marks the effect of the mean population growth rate in the data (2.5%).

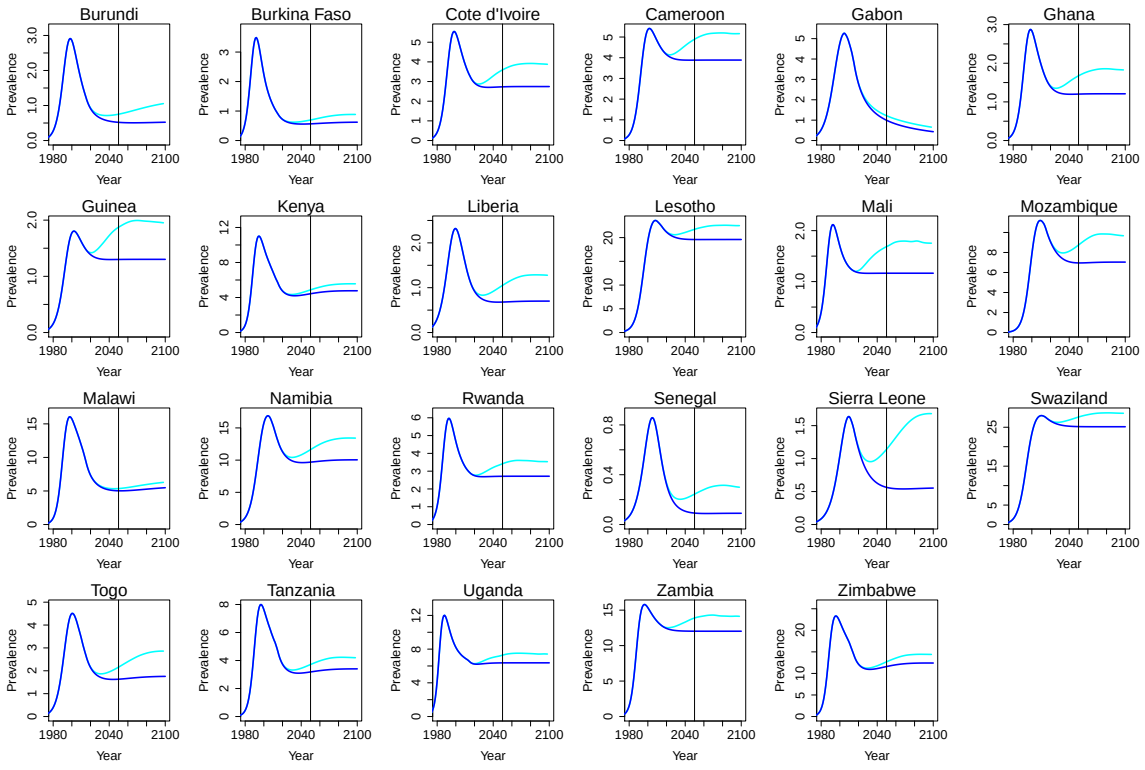


Figure A.8: The long-run effect of climate change on HIV prevalence under the RCP2.6 scenario (cyan line) relative to the no-warming scenario (blue line). Under an RCP2.6 scenario temperatures stabilize around 2050. However, the plots show that HIV rates continue to increase in most countries even after this date. The black vertical line highlights the year 2050.

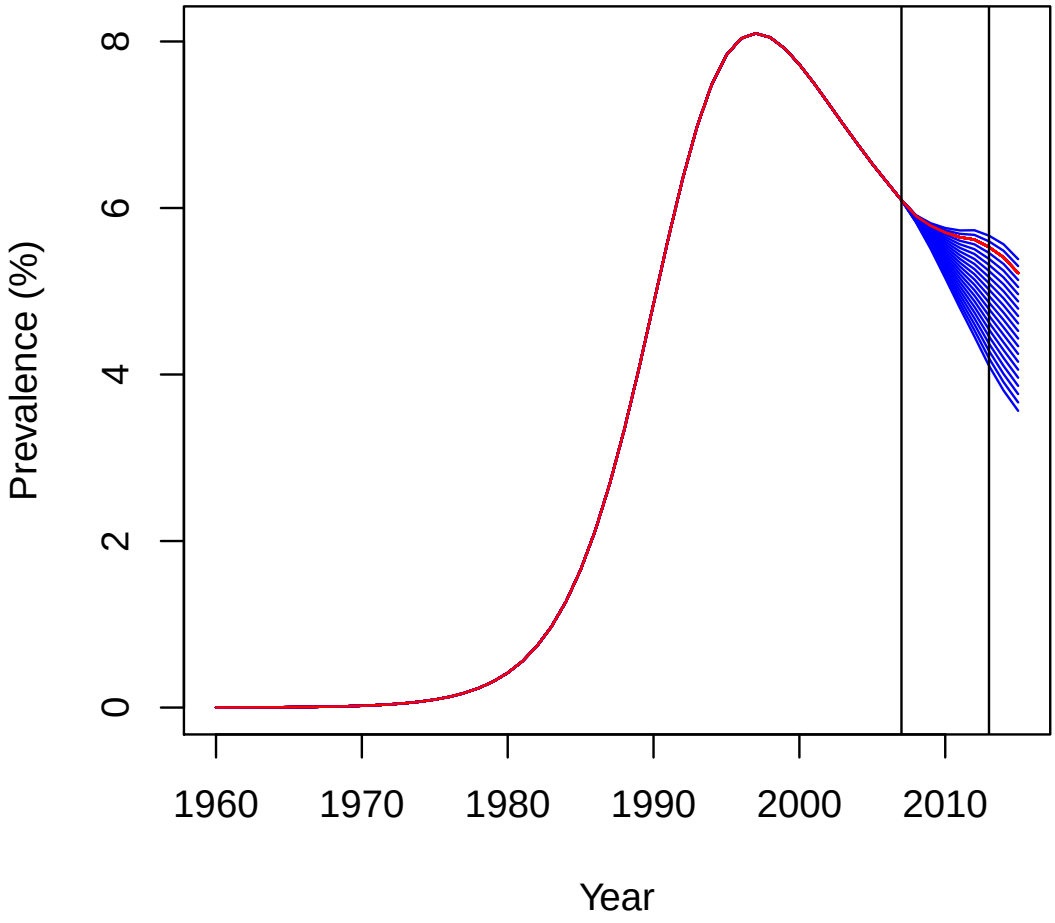


Figure A.9: A visual example of the process for estimating the relationship between temperature changes and p_H for Tanzania. Each blue line represents the effect of a different contact rate. The black vertical lines represents a six year period (the average length of time between two surveys). The red line shows the shift in the contact rate that best represents the shift in HIV prevalence predicted for this country, over this time period, for 1°C of warming.

Appendix B

Dynamic response of airborne infections to climate change: predictions for varicella

Absolute Humidity

The NARR dataset does not provide absolute humidity explicitly. I derive a conversion from relative humidity to absolute humidity using temperature. Relative humidity is defined as:

$$RH = (AH/SVD) * 100 \tag{B.1}$$

where AH is absolute humidity and SVD is saturation vapor density. The saturation vapor density of water is a nonlinear function of temperature that can be approximate in the 0-40 degree Celsius range with the following formula:

$$SVD = 5.018 + 0.32321T + 8.1847(10^{-3})(T^2) + 3.1243(10^{-4})(T^3) \tag{B.2}$$

where T is temperature in Celsius. Rearranging this gives a formula for absolute humidity which holds as long as temperature is in the 0-40°C range:

$$AH = (RH/100)5.018 + 0.32321T + 8.1847(10^{-3})(T^2) + 3.1243(10^{-4})(T^3) \quad \text{(B.3)}$$

I replace relative humidity with absolute humidity in the model. Results of this regression are shown in table B.1.

	Dependent variable:
	Log Transmission
Absolute Humidity	-0.049^{***}
	(0.016)
Temperature	0.011
	(0.023)
Observations	22,243
R^2	0.564
Adjusted R^2	0.537
Residual Std. Error	0.217 (df = 20961)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table B.1: Regression results showing the effect of absolute humidity on log transmission. Absolute humidity is significantly associated with an increase in incidence. However, using absolute humidity as opposed to relative humidity does not change the significance of temperature.

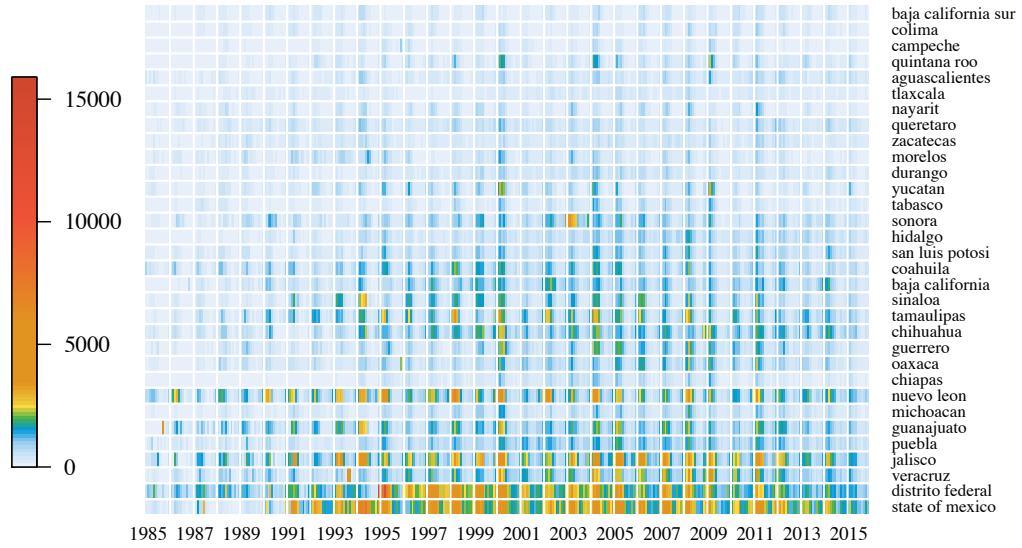


Figure B.1: Reported cases of varicella for each state in Mexico from 1985 -2016. Each row shows the time series of reported cases for a state. The states are arranged in ascending order of population size.

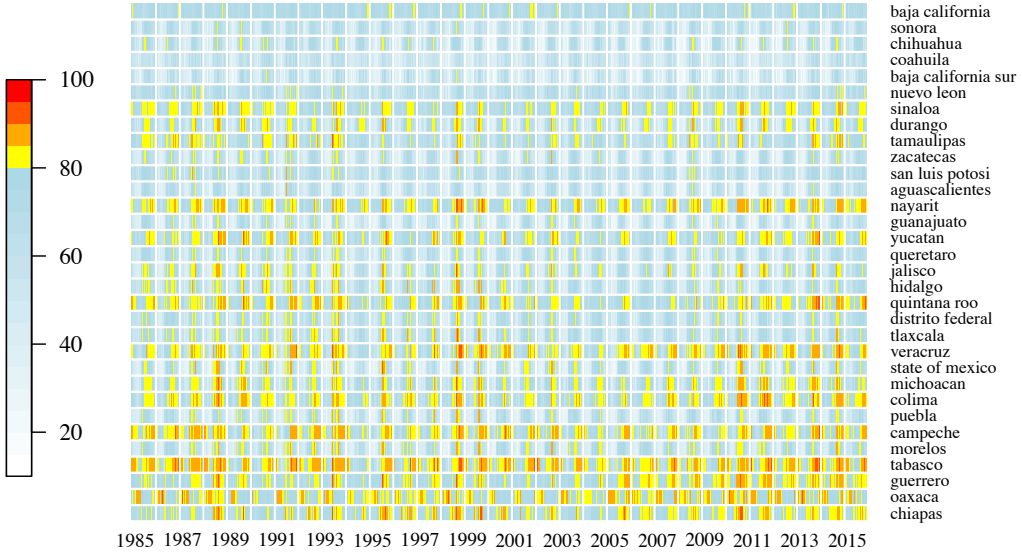


Figure B.2: Relative humidity (%) time series for each state in Mexico from 1985 -2016. Each row shows the humidity data for a state. The states are arranged according to latitude from north to south.

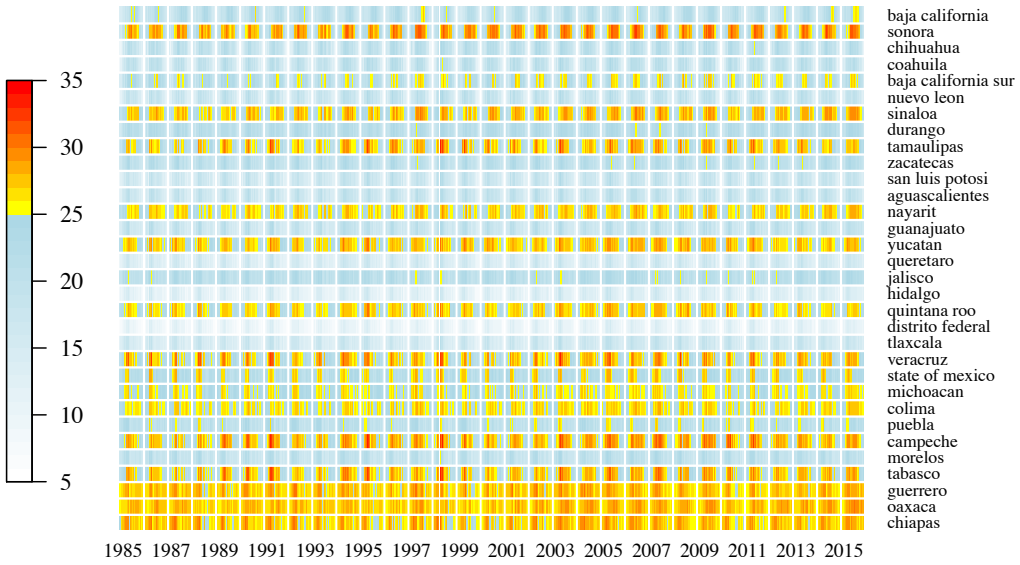


Figure B.3: Temperature time series (degree C) for each state in Mexico from 1985 -2016. Each row shows the temperature data for a state. The states are arranged according to latitude from north to south.

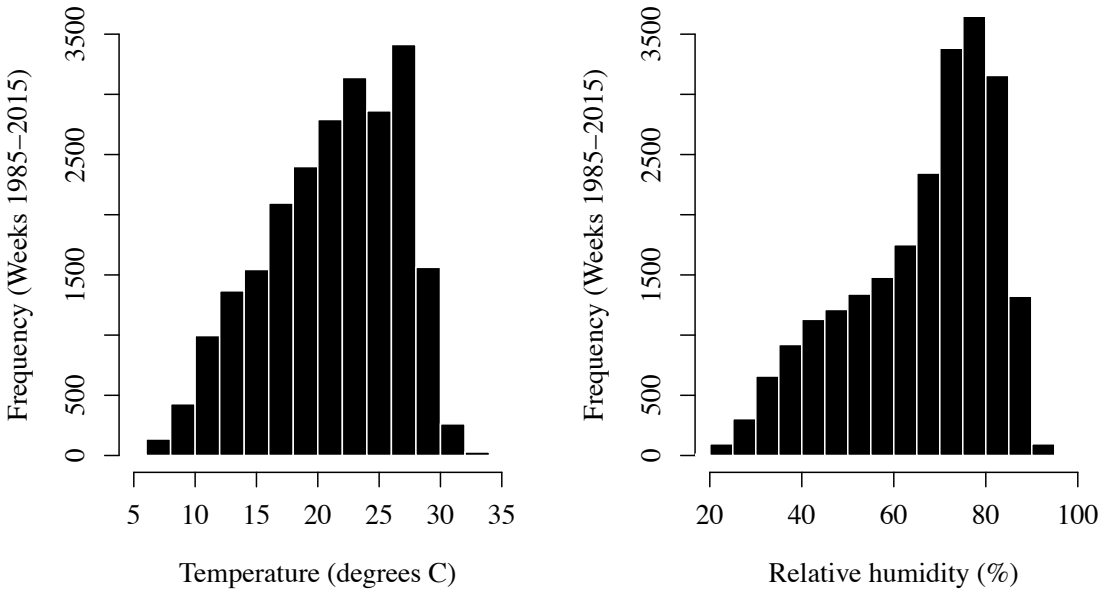


Figure B.4: The distributions of temperature and relative humidity across all states in Mexico between 1985 to 2015. The temperature range is 5°C to 35°C and the relative humidity range is 20% to 90%.

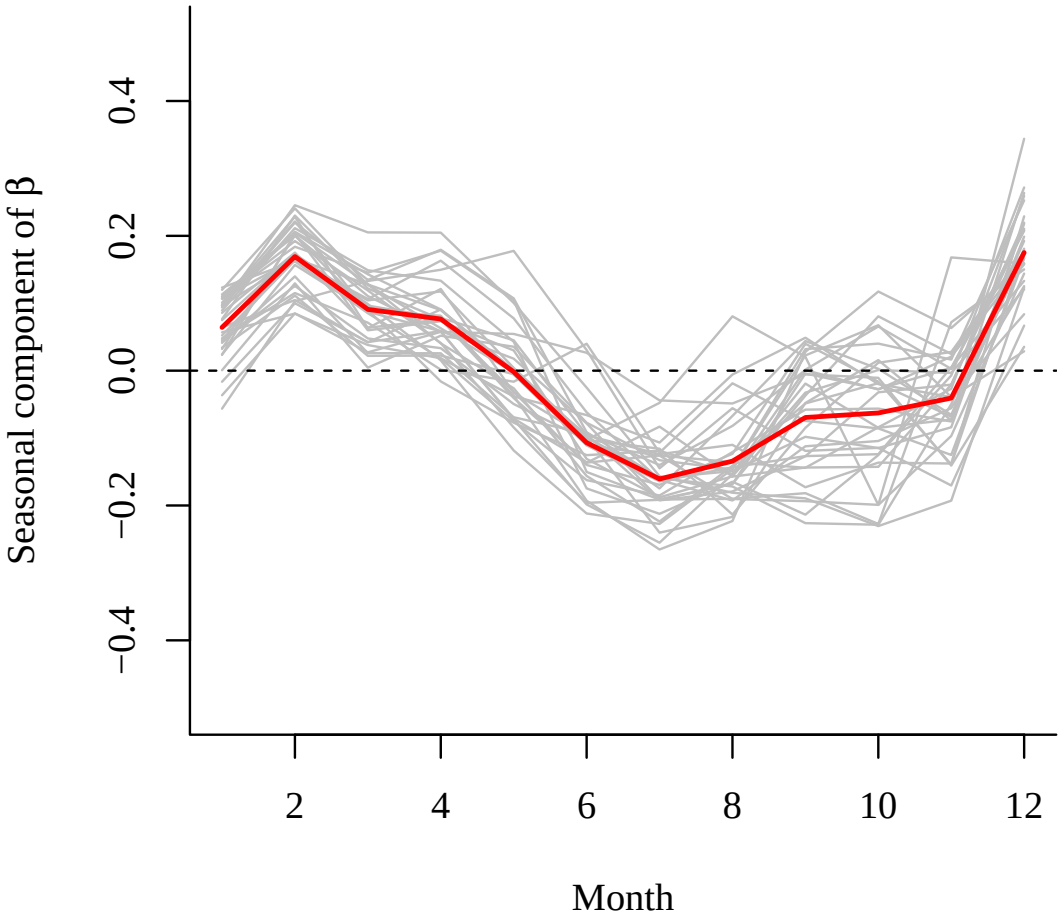


Figure B.5: Seasonal trend in transmission rates estimated by fitting the TSIR model, with environmental drivers and monthly factors, to each state. The grey lines indicate the mean-centered coefficients on the monthly factors for each state. The red line indicates the average across all states. Transmission is lower in the summer months, which corresponds to the summer school holidays in Mexico.

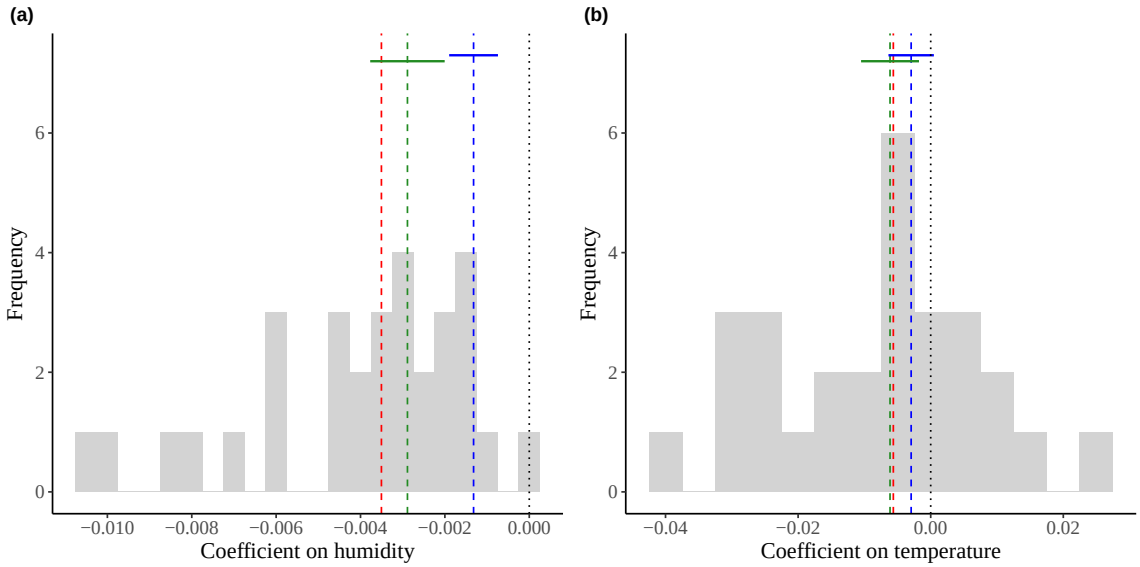


Figure B.6: Estimated effect of (a) humidity and (b) temperature on varicella transmission assuming an exponential relationship between environmental drivers and transmission. The histogram shows the distribution of coefficients on humidity and temperature from fitting the TSIR model to each state separately. Only coefficients that were statistically significant at the 95% level were included. The red dashed line indicates the median of the TSIR coefficients on humidity (-0.004, s.d. 0.0026) and temperature (-0.006, s.d. 0.0153). The vertical dashed green line indicates the estimated coefficient on humidity (-0.003; clustered s.e. 0.0005) and temperature (-0.006; clustered s.e. 0.0022) from the panel regression on reported cases. The horizontal solid green line indicates the 95% confidence interval. The vertical dashed blue line indicates the estimated coefficient on humidity (-0.001, clustered s.e. 0.0003) and temperature (-0.003; clustered s.e. 0.0017) from our combined approach (panel regression on empirical transmission rates). The horizontal solid blue line indicates the 95% confidence interval.

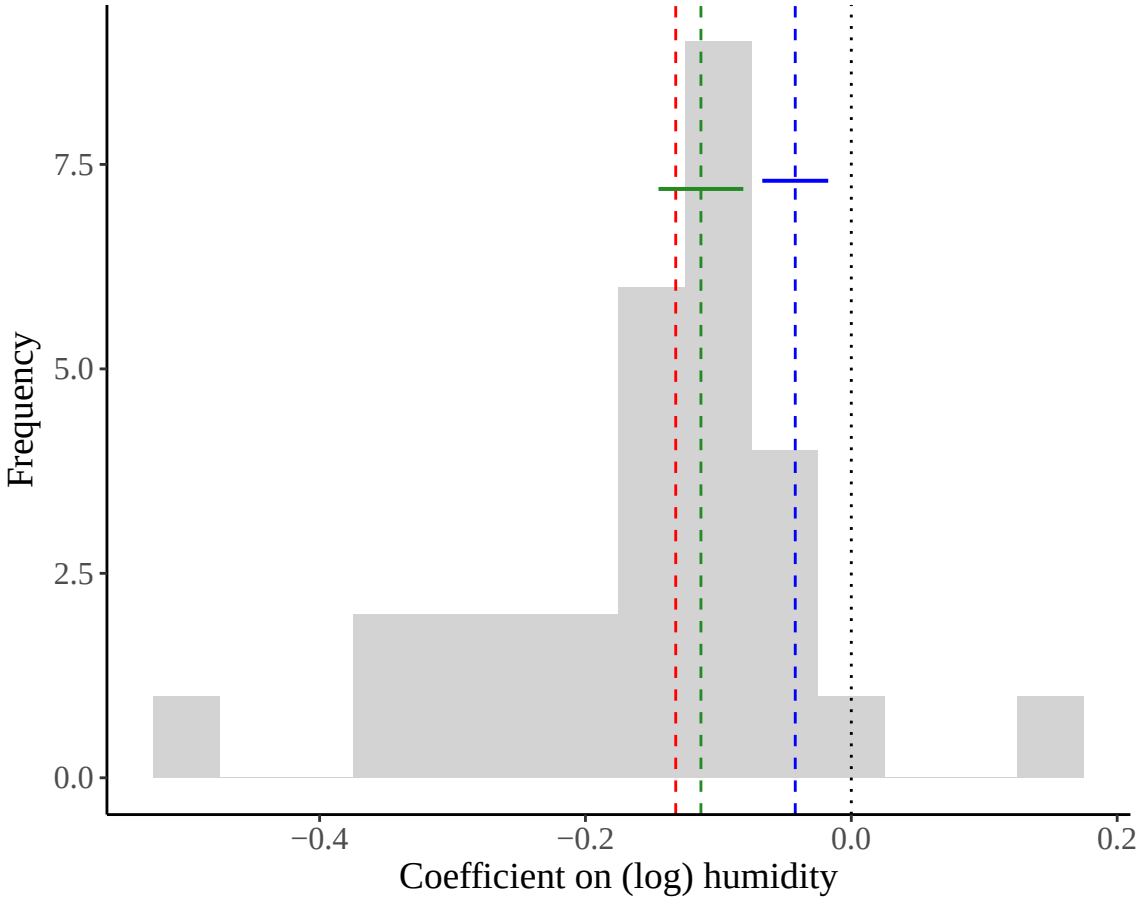


Figure B.7: Estimated effect of humidity on varicella transmission, estimated without including temperature in the model and assuming a log-linear relationship between humidity and transmission. The histogram shows the distribution of coefficients on log humidity from fitting the TSIR model to each state separately. Only coefficients that were statistically significant at the 95% level were included. The red dashed line indicates the median of the TSIR coefficients on humidity (-0.132; s.d. 0.1237). The vertical dashed green line indicates the estimated coefficient on log humidity (-0.113; clustered s.e. 0.0163) from the panel regression on reported cases. The horizontal solid green line indicates the 95% confidence interval. The vertical dashed blue line indicates the estimated coefficient on log humidity (-0.0422, clustered s.e. 0.0126) from our combined approach (panel regression on empirical transmission rates). The horizontal solid blue line indicates the 95% confidence interval.

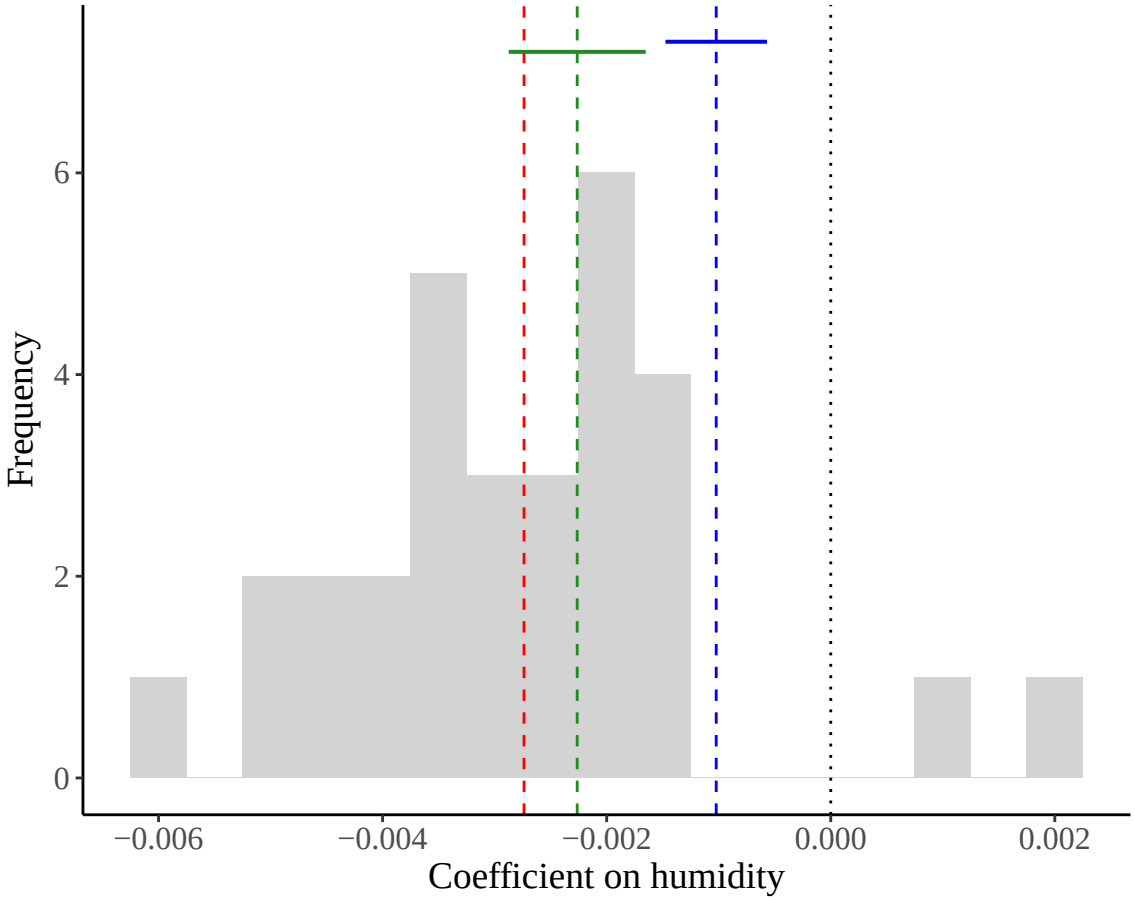


Figure B.8: Estimated effect of humidity on varicella transmission, estimated without including temperature in the model and assuming an exponential relationship between humidity and transmission. The histogram shows the distribution of coefficients on log humidity from fitting the TSIR model to each state separately. Only coefficients that were statistically significant at the 95% level were included. The red dashed line indicates the median of the TSIR coefficients on humidity (-0.003; s.d. 0.0017). The vertical dashed green line indicates the estimated coefficient on log humidity (-0.002; clustered s.e. 0.0003) from the panel regression on reported cases. The horizontal solid green line indicates the 95% confidence interval. The vertical dashed blue line indicates the estimated coefficient on log humidity (-0.001, clustered s.e. 0.0002) from our combined approach (panel regression on empirical transmission rates). The horizontal solid blue line indicates the 95% confidence interval.

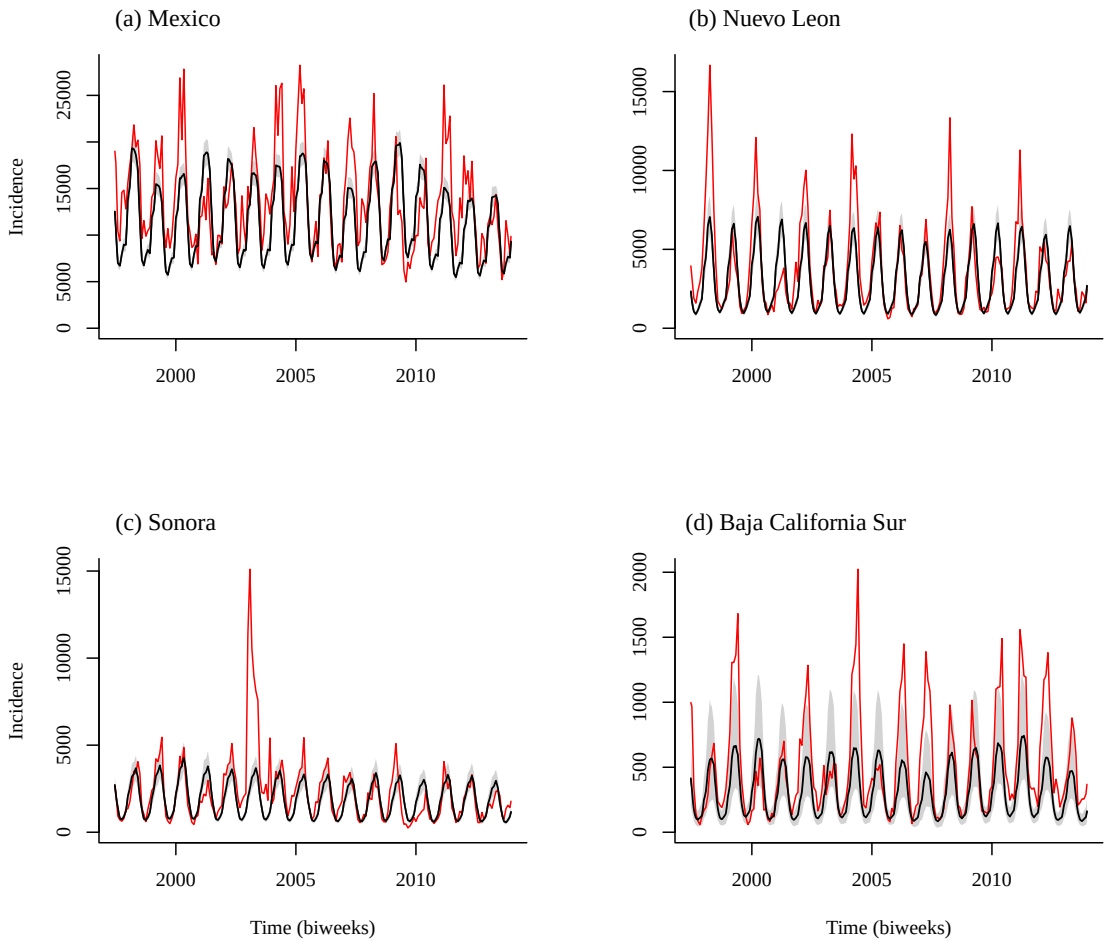


Figure B.9: Examples of simulated varicella incidence for four states in Mexico, with varying population sizes. Incidence was simulated using an SIR model with environmental drivers. Black line shows the median of 500 stochastic simulations using observed births and population size for each state, the temperature and humidity time series for each state, TSIR estimates for the monthly trend in the transmission rates, and the estimated effect of temperature and humidity on transmission. The shadowed grey region corresponds to the range between the 10th and 90th percentiles of the simulations. Red line shows the the reported varicella cases, corrected for underreporting.

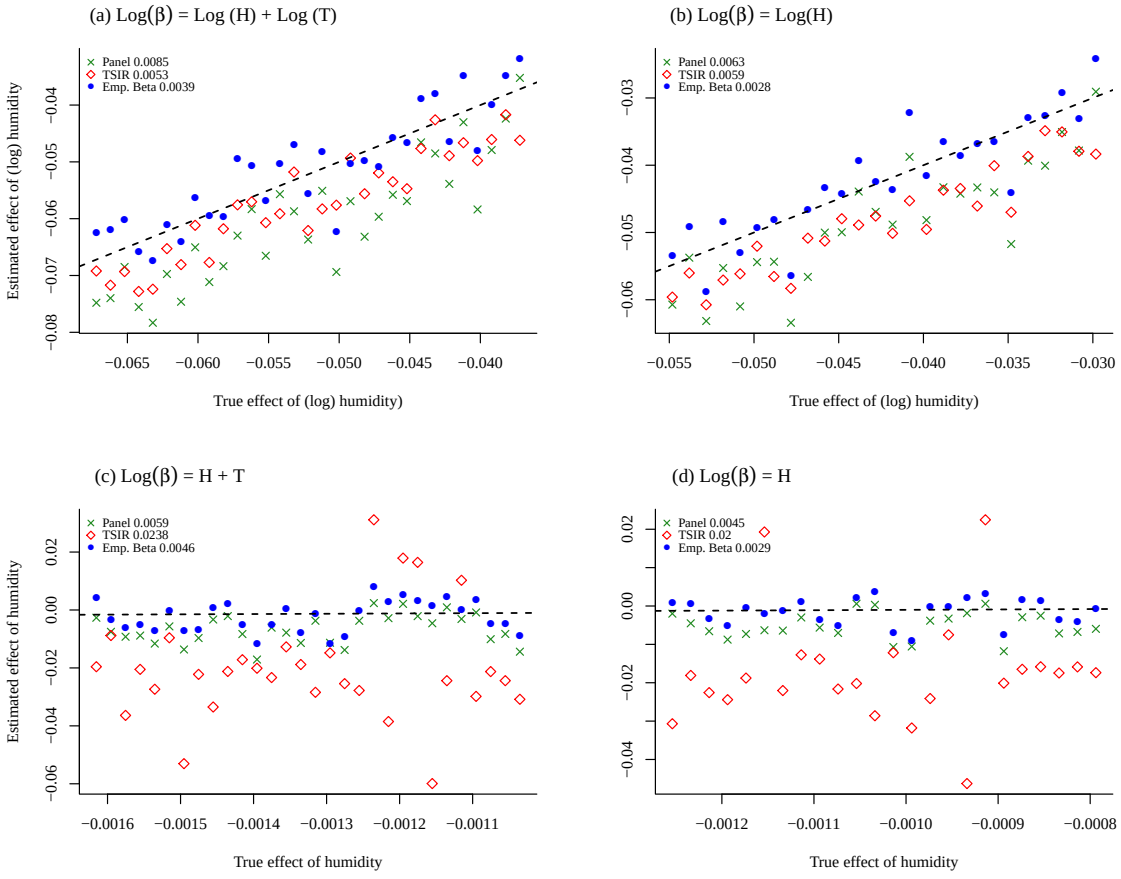


Figure B.10: Model robustness check on simulated incidence data, assuming four different functional forms for the relationship between varicella transmission and environmental drivers. Varicella incidence for each state in Mexico from 1985 to 2016 was simulated using an SIR model, with the observed births and population size, under four scenarios: (a) assuming a log-linear relationship between transmission and humidity (H) and temperature (T) (our main model); (b) assuming a log-linear relationship between transmission and humidity, excluding temperature; (c) assuming an exponential relationship between transmission and both environmental drivers; and (d) assuming an exponential relationship between transmission and humidity; excluding temperature. In each plot, the known effect of humidity on transmission is plotted against the estimated effect of humidity on transmission. The dashed black line is the x=y line. The green crosses represent the estimates from the panel regression on reported cases. The red diamonds represent the median of the estimates from the TSIR model fit to each state separately. The blue circles represent the estimates from the combined method using panel regression on empirical transmission rates. The mean absolute error for each method is on the top left corner.

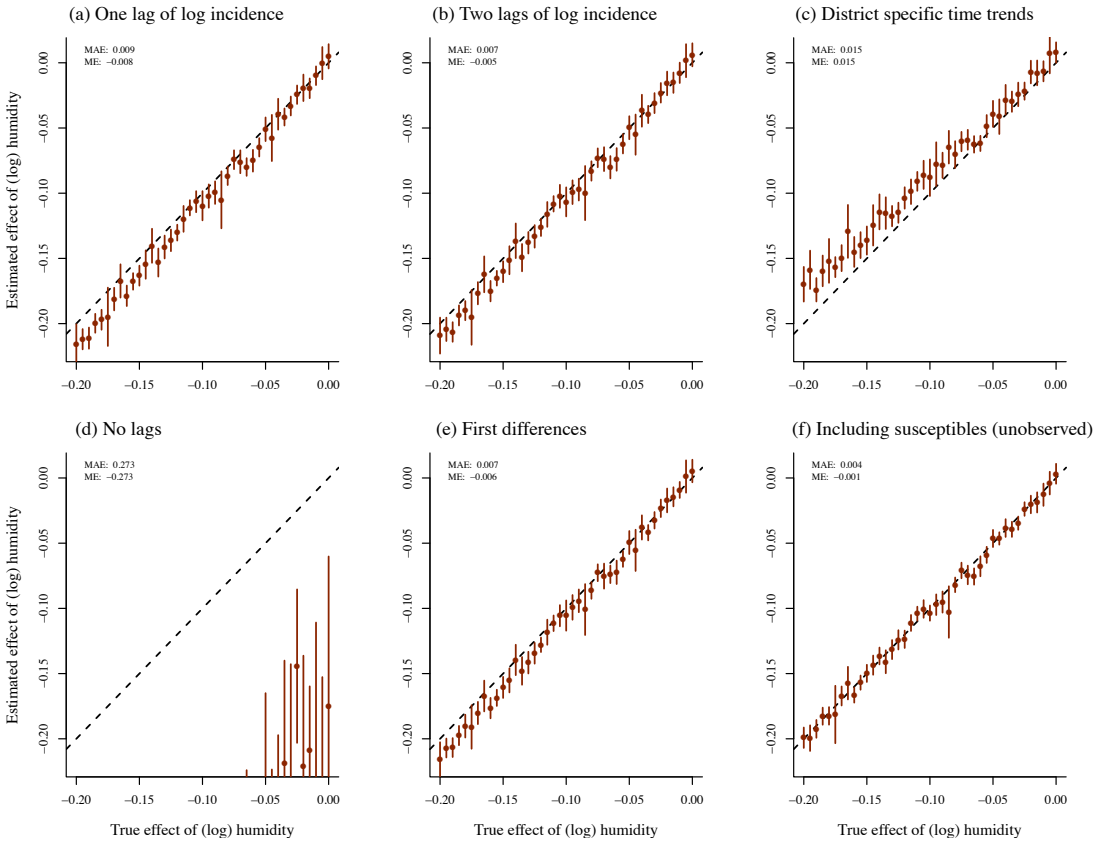


Figure B.11: Testing alternate specifications of the panel model and their accuracy at estimating the known effect of log humidity in our simulations. Varicella incidence for each state in Mexico from 1985-2016 was simulated using an SIR model, allowing the transmission rate to have a time trend (the time trend was obtained by fitting a cubic smoothing spline to the yearly trends in the transmission rate from the TSIR fit). In each plot, the known effect of humidity on transmission is plotted against the estimated effect of humidity on transmission; the dashed black line is the x=y line; the mean absolute error (MAE) and mean error (ME) is shown on the top right. (a) shows the results of using an AR(1) model including one lagged dependent variable (incidence). (b) shows the results of using an AR(2) model including the lagged dependent variable over two time steps. (a) and (b) show downward bias. We attempt to correct for this by replacing state-by-year fixed effects with state specific time trends. This causes a positive bias (c). We use a panel model with no lagged dependent variable as shown in (d) and also take first differences of incidence at time t+1 and incidence at time t (f). Finally we demonstrate that including the number of susceptibles explicitly in the model results in the lowest mean absolute error and mean error (e). This suggests the panel model is not able to accurately account for changes in the susceptible population over time.

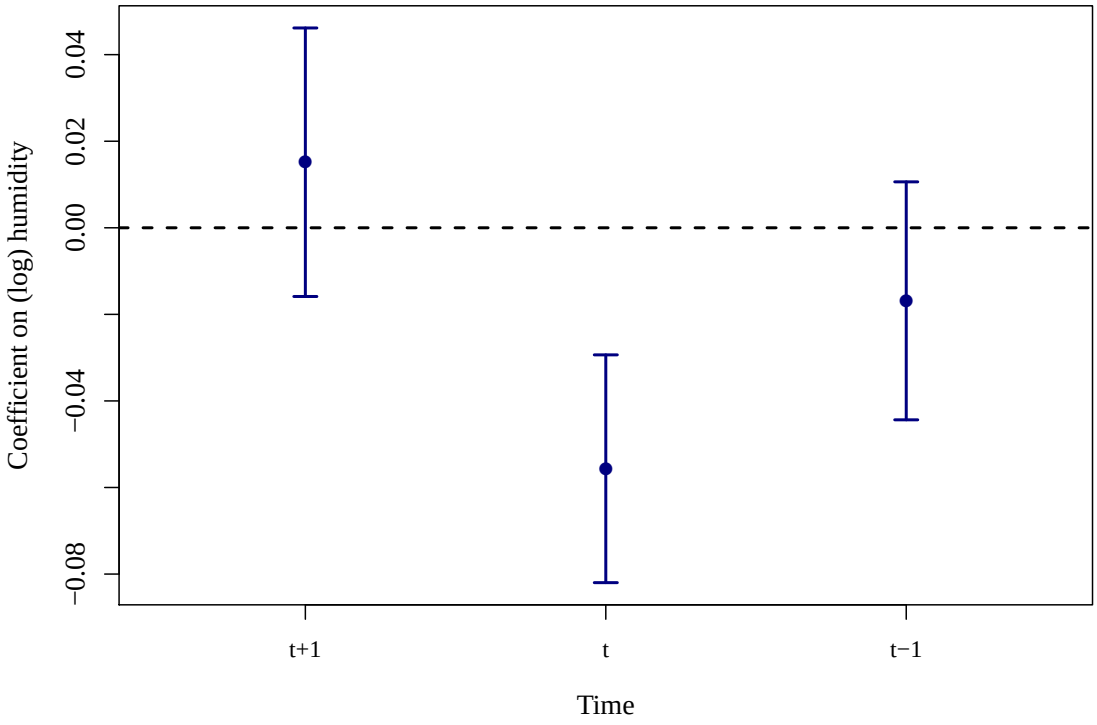


Figure B.12: The effect of humidity at different time steps on the transmission rate. Time (t) is indexed by time of disease transmission such that time (t+1) is two weeks after transmission occurs and time (t-1) is two weeks before the transmission occurs. Humidity at time t+1 and t-1 does not have a significant effect on transmission of varicella.

Appendix C

Implications of climate change for urban labor supply



Figure C.1: Data: the locations of the 6 cities used in the Pesquisa Mensal de Emprego employment survey.

Policy	Cooling Effect	Cost
Cool Roofs	0.2C (50% area)	7.5($/m^2$)
Green Roofs	0.45C (50% area)	161.5 ($/m^2$)
Cool Pavements	0.6C (100% area)	8.6 ($/m^2$)
Cool Roofs + Pavements	0.8C	7.5 ($/m^2$) + 8.6 ($/m^2$)

Table C.1: The four policies for reducing urban heating based on [37].

City	Pop 2014	Pop 2100
Recife	1,608,488	1,932,530
Salvador	2,902,927	3,487,744
Belo Horizonte	2,491,109	2,992,962
Rio de Janeiro	6,453,682	7,753,827
Sao Paulo	11,895,893	14,292,413
Porto Alegre	1,472,482	1,769,125

Table C.2: The population numbers from 2014 (column 1) projected to 2100 (column 2) based on World Bank data on urban population growth rates in Brazil. Growth rates post-2050 are extrapolated using a fitted polynomial. The minimum growth rate set to zero so that populations in urban areas do not decline.

City Names	Pay ($)	Air-Con(%)	TMax (°C)	TMax-UHI (°C)
Recife	4.37	6.26	30.31	30.11
Salvador	4.53	2.67	31.13	32.22
Belo Horizonte	5.97	1.63	27.39	30.02
Rio de Janeiro	6.35	24.46	28.75	33.62
Sao Paulo	6.21	3.40	26.87	31.31
Porto Alegre	5.87	9.38	24.65	27.81

Table C.3: Key figures for each city. The first column is mean pay in 2013 US dollars, used for projections. The second column is air-conditioning penetration rates for the state in which the city is located. This is based on a household survey from 2010. The third column gives average temperatures over the course of the survey, according to BEST data. The fourth column gives these temperatures using a simple adjustment for urban heat island effects.

VARIABLES	(1) Education	(2) Gender-Woman	(3) Age-(14>x>66)
Tmax Bin_0_15C	-0.00427***	0.000233*	0.0254***
	(0.000993)	(0.000107)	(0.00348)
Tmax Bin_15_18C	0.00166	0.000494*	0.00262
	(0.00103)	(0.000232)	(0.00424)
Tmax Bin_18_21C	-0.00118	-0.000478	0.00232
	(0.00172)	(0.000289)	(0.00268)
Tmax Bin_21_24C	-0.000872**	5.99e-05	0.00297
	(0.000302)	(0.000114)	(0.00626)
Tmax Bin_24_27C	0.000591	-0.000114	4.79e-05
	(0.00139)	(6.52e-05)	(0.00661)
Tmax Bin_30_33C	-0.000385	4.52e-05	0.000745
	(0.00106)	(0.000141)	(0.00450)
Tmax Bin_33_36C	0.000205	0.000134	-0.000378
	(0.00174)	(0.000182)	(0.00485)
Tmax Bin_36_40C	0.0139	0.00144	-0.0216***
	(0.0147)	(0.000908)	(0.00368)
Constant	3.489***	0.528***	34.67***
	(0.00445)	(0.000646)	(0.0203)
Observations	9,079,175	14,177,246	10,068,402
R-squared	0.014	0.000	0.007
YearXCity	Yes	Yes	Yes
CityXMoY	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C.4: The results of selection tests using the non-parametric binned model. If selection is not occurring then 1/20 results should be significant.

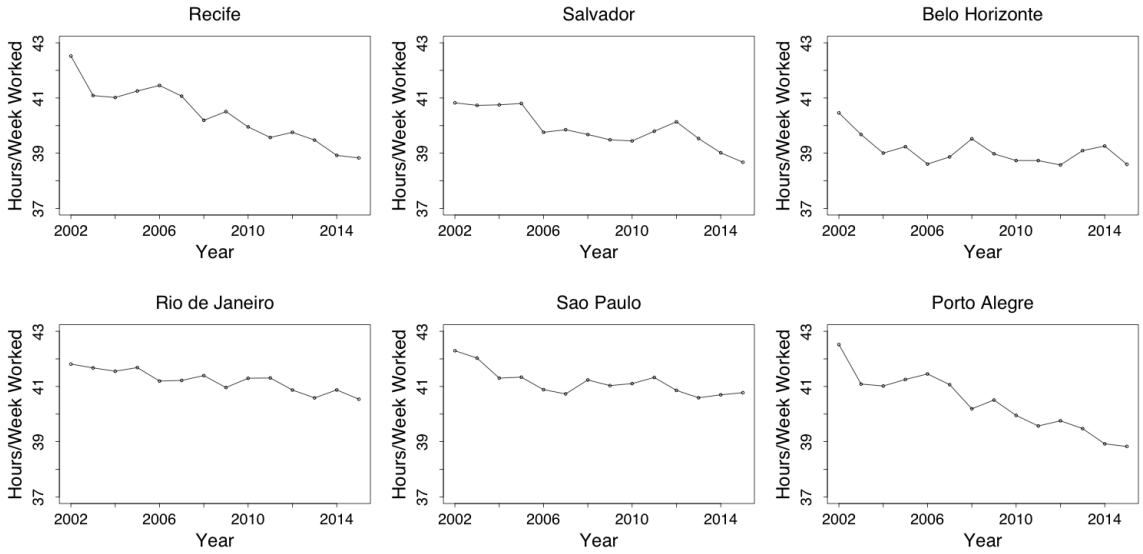


Figure C.2: Data: trends in working hours over time for each city.

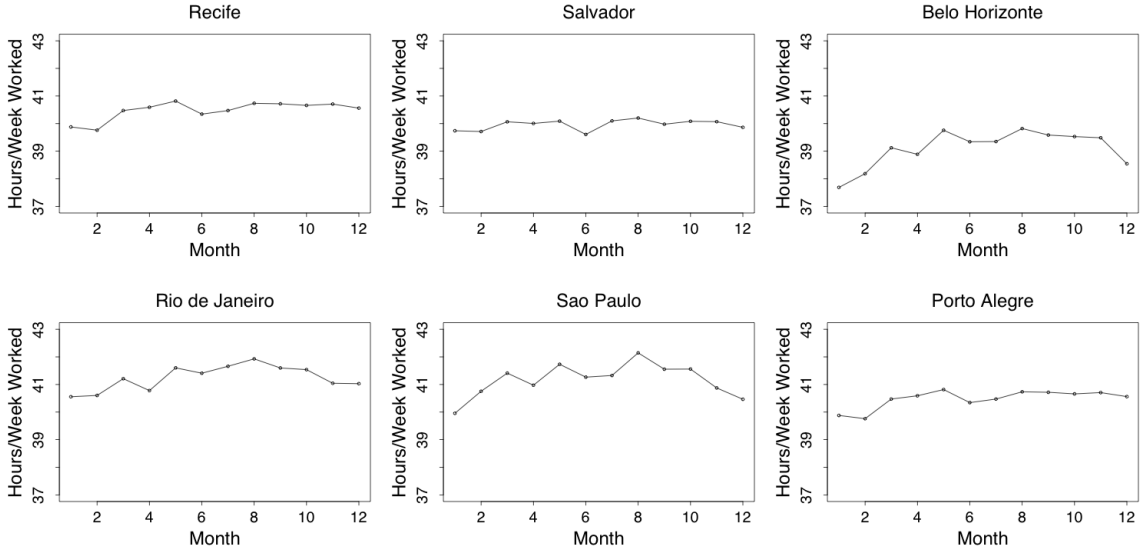


Figure C.3: Data: seasonal patterns of work for each city.

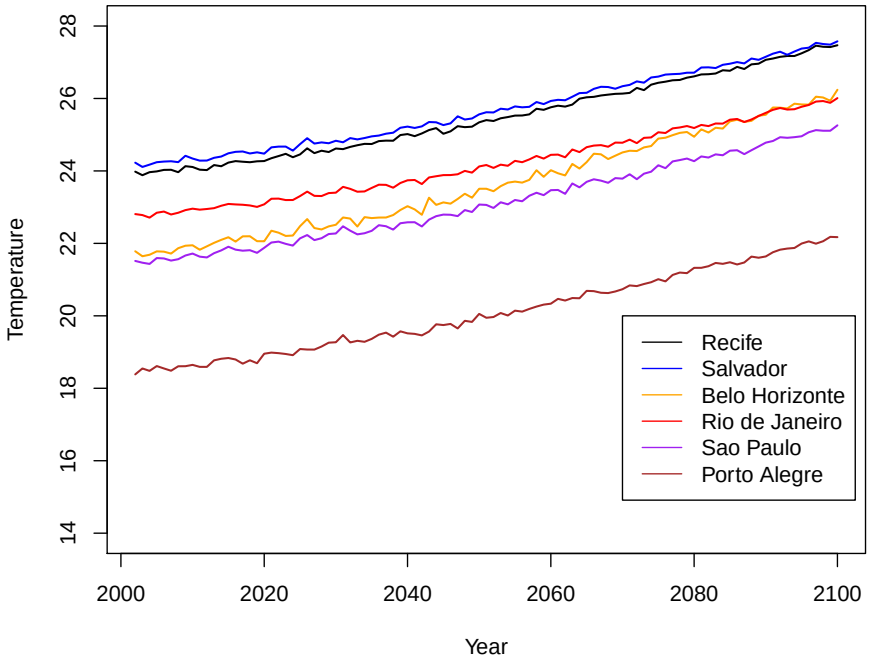


Figure C.4: Data: CMIP5 temperature projections for each city.

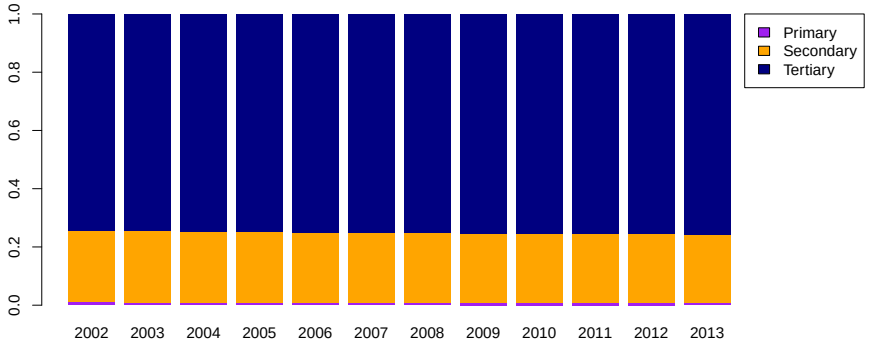


Figure C.5: The ratio of primary, secondary and tertiary sector employment in the sample for each year from 2002-2013. Ratios between sectors appear stable over time.

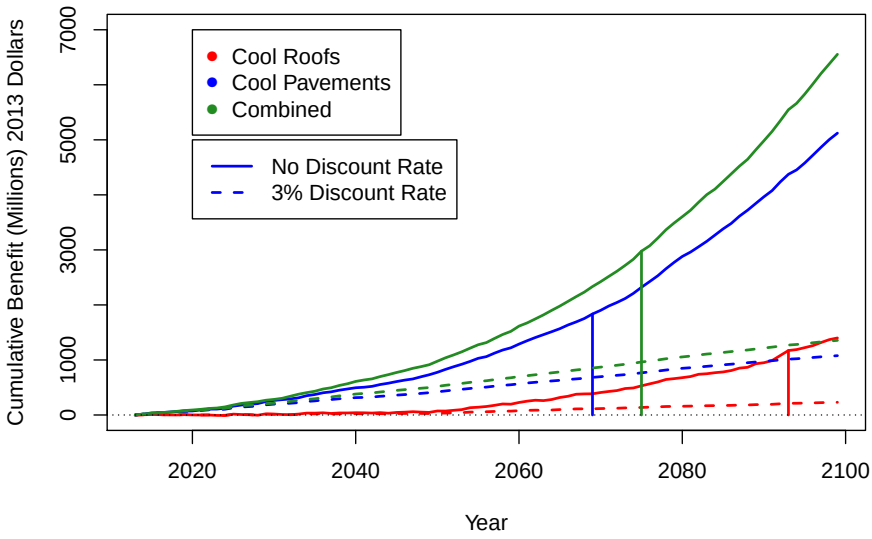


Figure C.6: The cumulative benefits accrued from implementing each of three policy options. Benefits are in terms of recovered damages from otherwise lost labor supply. The date at which these cumulative benefit equal the costs of implementing the policy is shown with the vertical lines. The cumulative benefits with an without discounting are shown. Applying a 3% discount rate means the benefits of cooling policies on labor supply do not outweigh the costs before 2100.

	Main Data	UHI Adjusted Data
Individual	0.000	0.000
City	0.000	0.000
Year, City, Month	0.016	0.001
Year∗City,Month∗City,Week	0.321	0.415

Table C.5: F-test results (Prob >F) for fourth order polynomial regressions run with different clusterings of standard errors. The results of main polynomial regression are robust for both the standard dataset and UHI adjusted dataset. Clustering errors at the Year*City level appears to reduce the significance of the F-test.

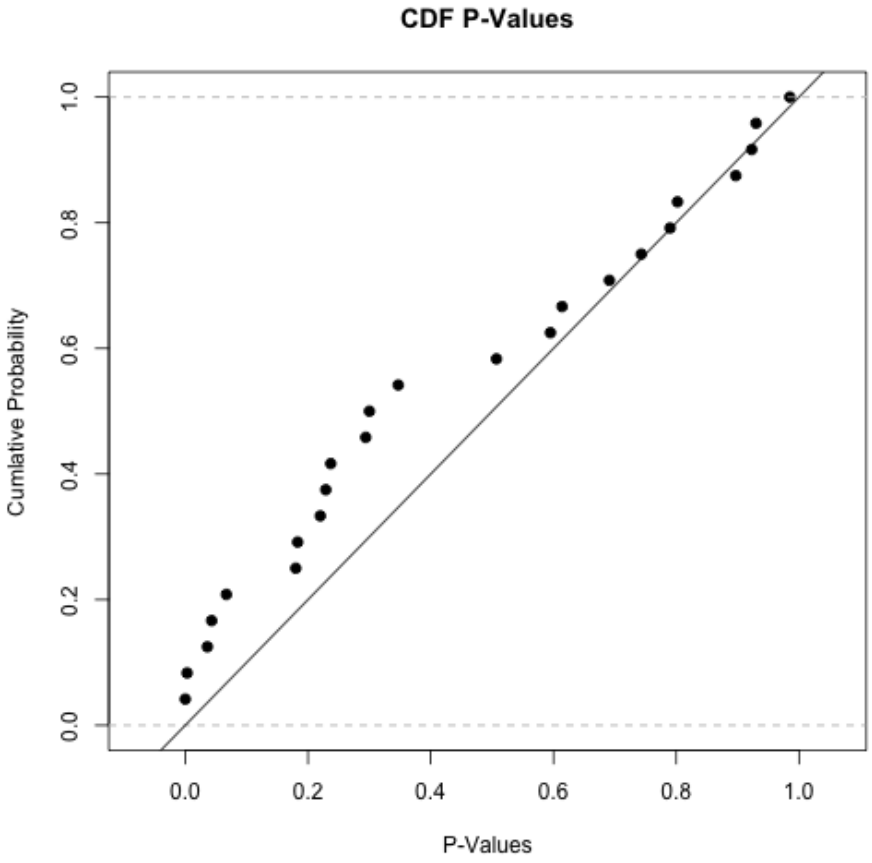


Figure C.7: P-values reported in the selection tests C.4.If there is no significant effect of temperature on selection then p-value should line along the x–y diagonal.

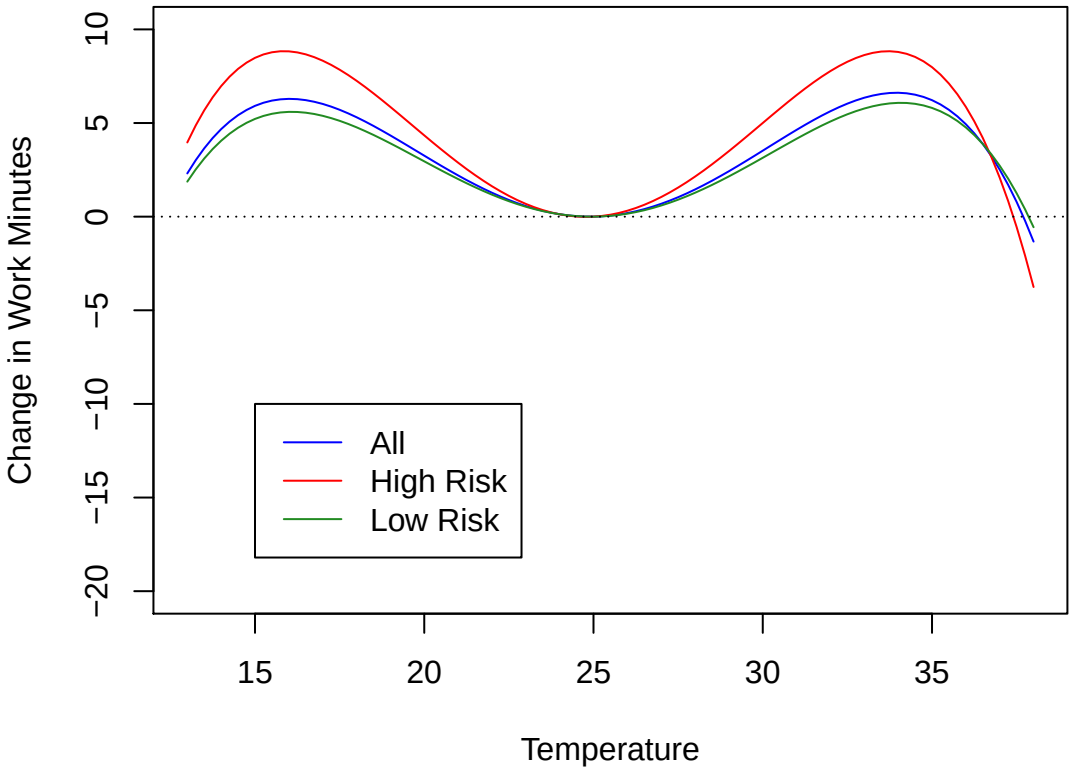


Figure C.8: Leads: The effect of temperature in the following week on labor supply in the current week.

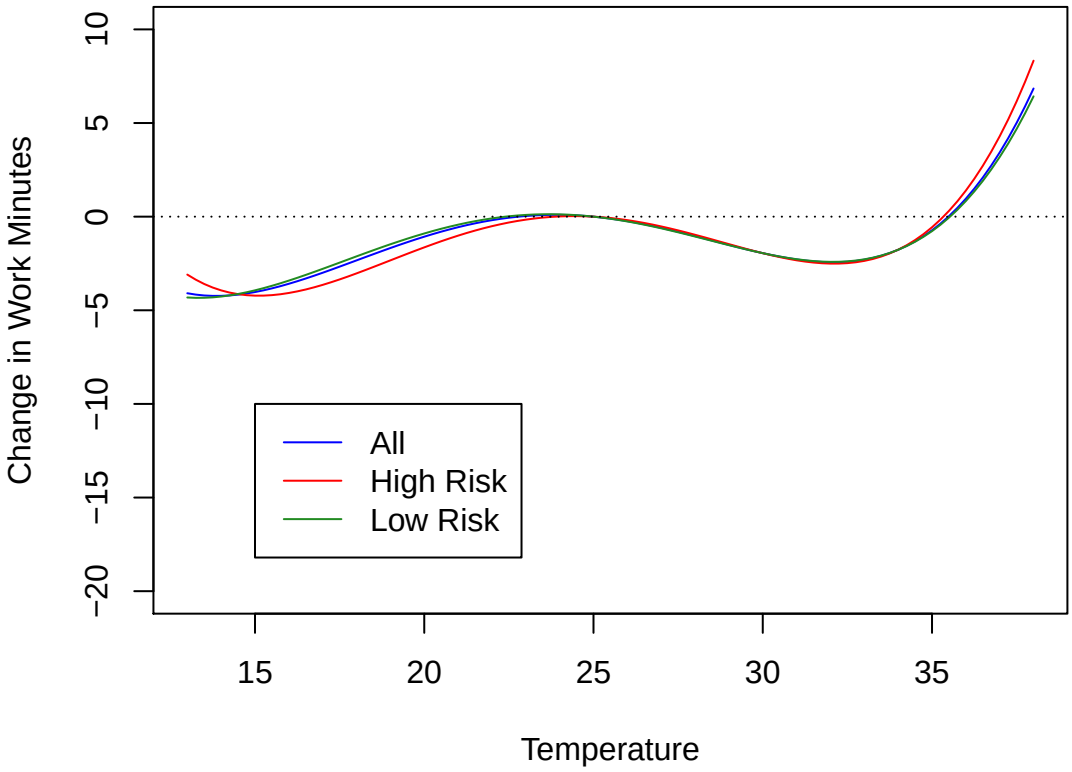


Figure C.9: Lags: The effect of temperature in previous week on labor supply in the current week.

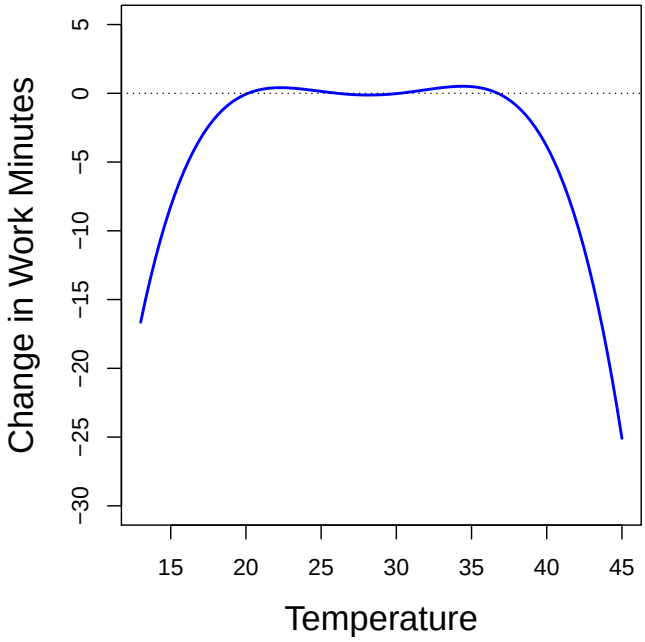


Figure C.10: The response function for labor supply with temperature data adjusted for urban heat islands.

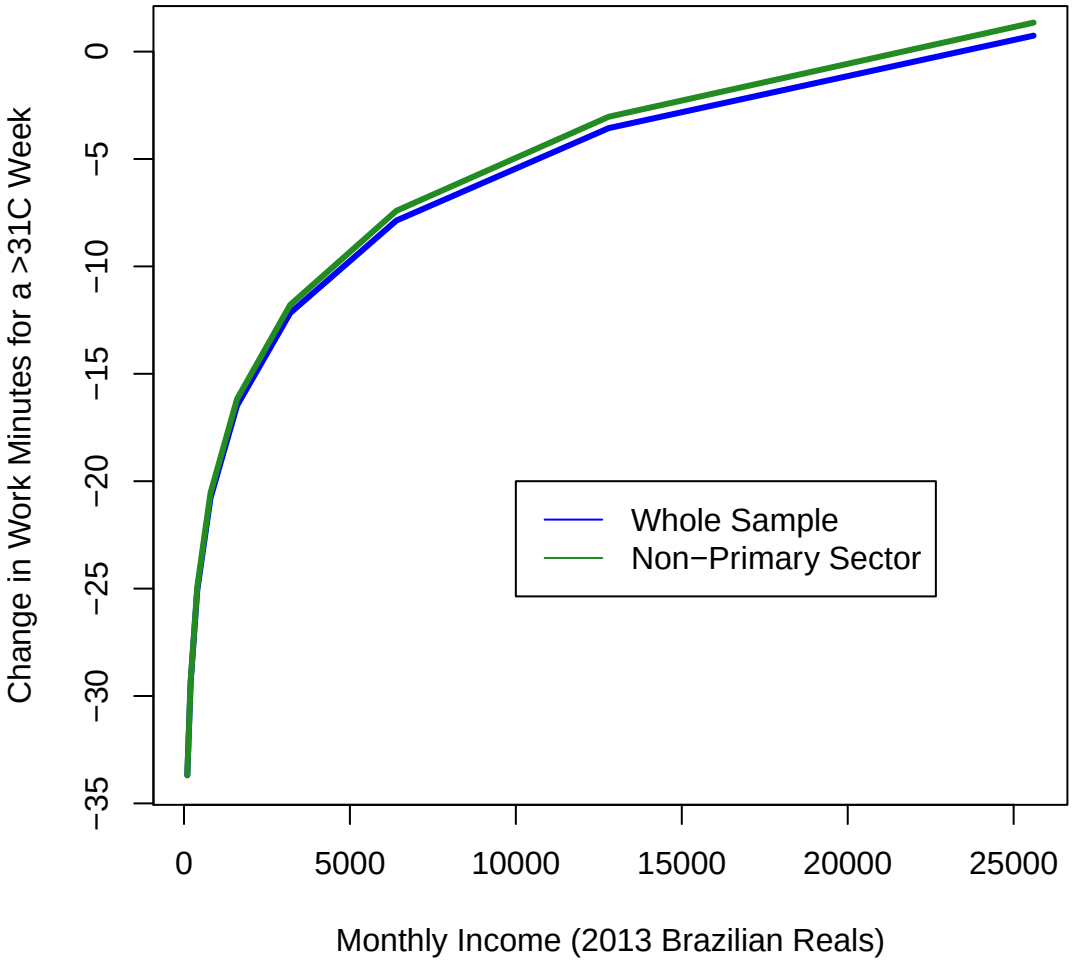


Figure C.11: The effect of economic income on the temperature-labor response was explored using interaction terms. First a new dummy variable was created for being in a week where temperatures were 31°C on average or higher. These weeks represent excessive hot weeks. This dummy was then interacted with log average monthly income, adjusted to 2013 Brazilian Reals. The results of the predicted effect of being in a hot week are plotted on the y-axis against unlogged income on the x-axis. The analysis was done of the whole sample (blue line) and a sub-sample of only non-primary sector employees (green line). The temperature term and the interaction term were significant: ($p<0.01$) and ($p<0.05$) respectively.

Bibliography

[1] Achyuta Adhvaryu, Namrata Kala, and Anant Nyshadham. The light and the heat: Productivity co-benefits of energy-saving technology. *Unpublished Manuscript*, 2014.

[2] Roy Anderson and Robert May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1991.

[3] Aranka Anema, Nicholas Vogenthaler, Edward A Frongillo, Suneetha Kadiyala, and Sheri D Weiser. Food insecurity and hiv/aids: current knowledge, gaps, and research priorities. *Current HIV/AIDS Reports*, 6(4):224–231, 2009.

[4] David Anthoff and Richard SJ Tol. The uncertainty about the social cost of carbon: A decomposition analysis using fund. *Climatic Change*, 117(3):515–530, 2013.

[5] WMO Regional Climate Centre at KNMI. CMIP5 coupled model intercomparison project.

[6] M. Auffhammer, S. M. Hsiang, W. Schlenker, and A. Sobel. Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, 7(2):181–198, 2013.

[7] R.E. Baker, A.L D'Agostino, T. Foreman, M. Greenstone, S.M. Hsiang, A. Hultgren, A.S Jina, and M Pecenco. Human productivity in a warmer world: The impact of climate change on the global workforce. *Working Paper*, 2017.

[8] The World Bank. World bank open data.

[9] Alan I Barreca and Jay P Shimshack. Absolute humidity, temperature, and influenza mortality: 30 years of county-level evidence from the United States. *American journal of epidemiology*, 176(suppl 7):S114–S122, 2012.

[10] VR Barros, CB Field, DJ Dokke, MD Mastrandrea, KJ Mach, TE Bilir, M Chatterjee, KL Ebi, YO Estrada, RC Genova, et al. Climate change 2014: impacts, adaptation, and vulnerability. Part B: regional aspects. contribution of Working Group ii to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. 2015.

[11] Ottar N. Bjørnstad, Brbel F. Finkenstädt, and Bryan T. Grenfell. Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs*, 72(2):169–184, 5 2002.

[12] Pratikshya Bohra-Mishra, Michael Oppenheimer, and Solomon M Hsiang. Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences*, 111(27):9780–9785, 2014.

[13] Susan M Bolton. *Interactions between HIV/AIDS and the environment: A review of the evidence and recommendations for next steps*. IUCN, 2010.

[14] Martin Brockerhoff and Ann E Biddlecom. Migration, sexual behavior and the risk of HIV in Kenya. *International migration review*, pages 833–856, 1999.

[15] H. Buhaug. Climate not to blame for African civil wars. *Proceedings of the National Academy of Sciences*, 107(38):16477–16482, 2010.

[16] Halvard Buhaug. Concealing agreements over climate-conflict results. *Pnas*, 111(6):E636, 2014.

[17] M Burke, SM Hsiang, and E Miguel. Global non-linear effect of temperature on economic production. *Nature*, 2015.

[18] Marshall Burke, Erick Gong, and Kelly Jones. Income shocks and hiv in africa. *The Economic Journal*, 2014.

[19] Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[20] Tamma A Carleton. Crop-damaging temperatures increase suicide rates in India. *Proceedings of the National Academy of Sciences*, page 201701354, 2017.

[21] Tamma A Carleton and Solomon M Hsiang. Social and economic impacts of climate. *Science*, 353(6304):aad9837, 2016.

[22] Climate Change. *Building a low-carbon economy ? the UK ' s contribution to tackling climate change*, volume 8. 2008.

[23] Myron S. Cohen, Ying Q. Chen, Marybeth McCauley, Theresa Gamble, Mina C. Hosseinipour, Nagalingeswaran Kumarasamy, James G. Hakim, Johnstone Kumwenda, Beatriz Grinsztejn, Jose H.S. Pilotto, Sheela V. Godbole, Suwat Chariyalertsak, Breno R. Santos, Kenneth H. Mayer, Irving F. Hoffman, Susan H. Eshleman, Estelle Piwowar-Manning, Leslie Cottle, Xinyi C. Zhang, Joseph Makhema, Lisa A. Mills, Ravindre Panchia, Sharlaa Faesen, Joseph Eron, Joel Gallant, Diane Havlir, Susan Swindells, Vanessa Elharrar, David Burns, Taha E. Taha, Karin Nielsen-Saines, David D. Celentano, Max Essex,

Sarah E. Hudelson, Andrew D. Redd, and Thomas R. Fleming. Antiretroviral therapy for the prevention of HIV-1 transmission. *New England Journal of Medicine*, 375(9):830–839, sep 2016.

[24] Marie Connolly. Here comes the rain again: Weather and the intertemporal substitution of leisure. *Journal of Labor Economics*, 26(1):73–100, 2008.

[25] Alex De Waal and Alan Whiteside. New variant famine: AIDS and food crisis in Southern Africa. *The Lancet*, 362(9391):1234–1237, 2003.

[26] Melissa Dell, Benjamin F Jones, and Benjamin a Olken. Climate shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, 4(3):66ß95, 2012.

[27] Melissa Dell, Benjamin F Jones, and Benjamin A Olken. What do we learn from the weather? the new climate–economy literature. *Journal of Economic Literature*, 52(3):740–798, 2014.

[28] Department of Energy & Climate Change. Carbon valuation in UK policy sppraisal: A revised approach. (July):1–128, 2009.

[29] T Deryugina and SM Hsiang. Does the environment still matter? Daily temperature and income in the United States. 2014.

[30] Olivier Deschenes and Enrico Moretti. Extreme weather events, mortality, and migration. *The Review of Economics and Statistics*, 91(4):659–681, 2009.

[31] Ethan R. Deyle, M. Cyrus Maher, Ryan D. Hernandez, Sanjay Basu, and George Sugihara. Global environmental drivers of influenza. *Proceedings of the National Academy of Sciences*, 113(46):13081–13086, 2016.

[32] Delavane B Diaz. Evaluating the key drivers of the US government's social cost of carbon: A model diagnostic and inter-comparison study of climate impacts in DICE, FUND, and PAGE. *Working Paper*, 2014.

[33] Kristin L Dunkle, Rachel K Jewkes, Heather C Brown, Glenda E Gray, James A McIntryre, and Sio?án D Harlow. Transactional sex among women in Soweto, South Africa: prevalence, risk factors and association with HIV infection. *Social science & medicine*, 59(8):1581–1592, 2004.

[34] Robert Engelman. The state of world population 2009. facing a changing world: Women, population and climate. In *The state of world population 2009. Facing a changing world: Women, population and climate*. UNFPA, 2009.

[35] U S Epa. Reducing Urban Heat Islands: Compendium of Strategies. *Heat Island Reduction Activities*, pages 1–23, 2008.

[36] F Estrada, WJW Botzen, and RSJ Tol. A global economic assessment of city policies to reduce climate change impacts. *Nature Climate Change*, 7, 2017.

[37] Francisco Estrada, W. J. Wouter Botzen, and Richard S. J. Tol. A global economic assessment of city policies to reduce climate change impacts. *Nature Climate Change*, 7(6):403–406, 2017.

[38] B. F. Finkenstädt and B. T. Grenfell. Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(2):187–205, 2000.

[39] Center for International Earth Science Information Network CIESIN Columbia University. Global urban heat island (UHI) data set, 2013.

[40] A. Fouillet, G. Rey, V. Wagner, K. Laaidi, P. Empereur-Bissonnet, A. Le Tertre, P. Frayssinet, P. Bessemoulin, F. Laurent, P. De Crouy-Chanel, E. Jougla, and Denis Hémon. Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003? A study of the 2006 heat wave. *International Journal of Epidemiology*, 37(2):309–317, 2008.

[41] Howard Frumkin, Jeremy Hess, George Luber, Josephine Malilay, and Michael McGeehin. Climate change: the public health response. *American Journal of Public Health*, 98(3):435–445, 2008.

[42] G P Garnett, M J Cox, D A Bundy, J M Didier, and J St Catharine. The age of infection with varicella-zoster virus in St Lucia, West Indies. *Epidemiology and infection*, 110(2):361–72, 1993.

[43] G P Garnett and B T Grenfell. The epidemiology of varicella-zoster virus infections : A mathematical model. *Epidemiology and infection*, 108(3):495–511, 1992.

[44] LM Gartland. *Heat islands: understanding and mitigating heat in urban areas*. 2012.

[45] Antonio Gasparrini, Yuming Guo, Masahiro Hashizume, Eric Lavigne, Antonella Zanobetti, Joel Schwartz, Aurelio Tobias, Shilu Tong, Joacim Rocklöv, Bertil Forsberg, et al. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*, 386(9991):369–375, 2015.

[46] Patrick Gerland, Adrian E Raftery, Hana Ševčíková, Nan Li, Danan Gu, Thomas Spoorenberg, Leontine Alkema, Bailey K Fosdick, Jennifer Chunn, Nevena Lalic, et al. World population stabilization unlikely this century. *Science*, 346(6206):234–237, 2014.

[47] Peter W Gething, Anand P Patil, David L Smith, Carlos A Guerra, Iqbal RF Elyazar, Geoffrey L Johnston, Andrew J Tatem, and Simon I Hay. A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malaria journal*, 10(1):378, 2011.

[48] Stuart Gillespie and Suneetha Kadiyala. *HIV/AIDS and food and nutrition security: From evidence to action*, volume 7. Intl Food Policy Res Inst, 2005.

[49] Kenneth Gillingham, William D Nordhaus, David Anthoff, Geoffrey Blanford, Valentina Bosetti, Peter Christensen, Haewon McJeon, John Reilly, and Paul Sztorc. Modeling uncertainty in climate change: A multi-model comparison. Technical report, National Bureau of Economic Research, 2015.

[50] Rene Gommes, Jacques du Guerny, Michael H Glantz, and Lee-Nah Hsu. Climate and HIV/AIDS: a hotspots analysis for early warning rapid response systems. 2004.

[51] Clark Gray and Valerie Mueller. Drought and population mobility in rural Ethiopia. *World development*, 40(1):134–145, 2012.

[52] Clark L Gray and Valerie Mueller. Natural disasters and population mobility in bangladesh. *Proceedings of the National Academy of Sciences*, 109(16):6000–6005, 2012.

[53] Michael Greenstone, Elizabeth Kopits, and Ann Wolverton. Developing a social cost of carbon for us regulatory analysis: A methodology and interpretation. *Review of Environmental Economics and Policy*, 7(1):23–46, 2013.

[54] WF Grether. Human performance at elevated environmental temperatures. *Aerospace Medicine*, 1973.

[55] John R Harris and Michael P Todaro. Migration, unemployment and development: a two-sector analysis. *The American economic review*, 60(1):126–142, 1970.

[56] Zeke Hausfather, Matthew J. Menne, Claude N. Williams, Troy Masters, Ronald Broberg, and David Jones. Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records. *Journal of Geophysical Research: Atmospheres*, 118(2):481–494, 1 2013.

[57] Simon I Hay, Dylan B George, Catherine L Moyes, and John S Brownstein. Big data opportunities for global infectious disease surveillance. *PLoS medicine*, 10(4):e1001413, 2013.

[58] Stephanie C Herring, Martin P Hoerling, James P Kossin, Thomas C Peterson, and Peter A Stott. Explaining extreme events of 2014 from a climate perspective. *Bulletin of the American Meteorological Society*, 96(12):S1–S172, 2015.

[59] Stephanie C Herring, Martin P Hoerling, Thomas C Peterson, and Peter A Stott. Explaining extreme events of 2013 from a climate perspective. *Bulletin of the American Meteorological Society*, 95(9):S1–S104, 2014.

[60] Herbert W Hethcote and James Yorke. *Gonorrhea transmission dynamics and control*, volume 56. Springer, 2014.

[61] Chris Hope. The marginal impact of CO2 from PAGE2002: An integrated assessment model incorporating the IPCC's five reasons for concern. *Integrated assessment*, 6(1), 2006.

[62] Solomon Hsiang. Climate econometrics. *Annual Review of Resource Economics*, 8:43–75, 2016.

[63] Solomon M Hsiang. Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of Sciences of the United States of America*, 107(35):15367–72, 2010.

[64] Solomon M Hsiang, Marshall Burke, and Edward Miguel. Quantifying the influence of climate on human conflict. *Science*, 341(6151):1235367, 2013.

[65] Solomon M. Hsiang and Kyle C. Meng. Reconciling disagreement over climate?conflict results in Africa. *Proceedings of the National Academy of Sciences*, 111(6):2100–2103, 2014.

[66] Marc L. Imhoff, Ping Zhang, Robert E. Wolfe, and Lahouari Bounoua. Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sensing of Environment*, 114(3):504–513, 3 2010.

[67] Seema Jayachandran. Air quality and early-life mortality evidence from indonesia's wildfires. *Journal of Human resources*, 44(4):916–954, 2009.

[68] Karen Jochelson, Monyaola Mothibeli, and Jean-Patrick Leger. Human immunodeficiency virus and migrant labor in south africa. *International Journal of Health Services*, 21(1):157–173, 1991.

[69] C. D. Kelly, D. L., & Kolstad. Integrated assessment models for climate change control. *International yearbook of environmental and resource economics*, (November 1998):171–197, 2000.

[70] Tord Kjellstrom, R Sari Kovats, Simon J Lloyd, Tom Holt, and Richard S J Tol. The direct impact of climate change on regional labor productivity. *Archives of Environmental & Occupational Health*, 64(4):217–227, 2009.

[71] Reto Knutti, Reinhard Furrer, Claudia Tebaldi, Jan Cermak, and Gerald A Meehl. Challenges in combining projections from multiple climate models. *Journal of Climate*, 23(10):2739–2758, 2010.

[72] Reto Knutti and Jan Sedláček. Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, 3(4):369–373, 2013.

[73] Suzanne Leclerc-Madlala. Age-disparate and intergenerational sex in southern africa: the dynamics of hypervulnerability. *Aids*, 22:S17–S25, 2008.

[74] Dan Li, Elie Bou-Zeid, Dan Li, and Elie Bou-Zeid. Synergistic interactions between urban heat islands and heat waves: The impact in cities is larger than the sum of its parts. *Journal of Applied Meteorology and Climatology*, 52(9):2051–2064, 9 2013.

[75] Ning Lin, Kerry Emanuel, Michael Oppenheimer, and Erik Vanmarcke. Physically based assessment of hurricane surge threat under climate change. *Nature Climate Change*, 2(6):462, 2012.

[76] David B Lobell, Marianne Bänziger, Cosmos Magorokosho, and Bindiganavile Vivek. Nonlinear heat effects on african maize as evidenced by historical yield trials. *Nature Climate Change*, 1(1):42–45, 2011.

[77] S Lolekha, W Tanthiphabha, P Sornchai, P Kosuwan, S Sutra, B Warachit, S Chup-Upprakarn, Y Hutagalung, J Weil, and H L Bock. Effect of climatic factors and population density on varicella zoster virus epidemiology within a tropical country. *The American Journal of Tropical Medicine and Hygiene*, 64(3):131–136, 2001.

[78] Anice C. Lowen, Samira Mubareka, John Steel, and Peter Palese. Influenza virus transmission Is dependent on relative humidity and temperature. *PLoS Pathogens*, 3(10):e151, 2007.

[79] Anice C Lowen and John Steel. Roles of humidity and temperature in shaping influenza seasonality. *Journal of virology*, 88(14):7692–5, 2014.

[80] A. S. Mahmud, C. J. E. Metcalf, and B. T. Grenfell. Comparative dynamics, seasonality in transmission, and predictability of childhood infections in Mexico. *Epidemiology and Infection*, 145(03):607–625, 2 2017.

[81] P H Mairiaux and Jacques Malchaire. Workers self-pacing in hot conditions: a case study. *Applied ergonomics*, 16(2):85–90, 1985.

[82] Luca Marchiori, Jean-François Maystadt, and Ingmar Schumacher. The impact of weather anomalies on migration in Sub-Saharan Africa. *Journal of Environmental Economics and Management*, 63(3):355–374, 2012.

[83] Douglas S Massey. *Strangers in a strange land: Humans in an urbanizing world*. WW Norton, 2005.

[84] Douglas S Massey, Joaquin Arango, Graeme Hugo, Ali Kouaouci, and Adela Pellegrino. *Worlds in Motion: Understanding International Migration at the End of the Millennium: Understanding International Migration at the End of the Millennium*. Clarendon Press, 1999.

[85] G.C. McCord. Malaria ecology and climate change. *The European Physical Journal Special Topics*, 225(3):459–470, 5 2016.

[86] Gerald A Meehl and Claudia Tebaldi. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, 305(5686):994–997, 2004.

[87] C Jessica E Metcalf, Ottar N Bjørnstad, Bryan T Grenfell, and Viggo Andreasen. Seasonality and comparative dynamics of six childhood infections in pre-vaccination Copenhagen. *Proceedings. Biological sciences / The Royal Society*, 276(1676):4111–8, 12 2009.

[88] Cari L Miller, David R Bangsberg, David M Tuller, Jude Senkungu, Annet Kawuma, Edward A Frongillo, and Sheri D Weiser. Food insecurity and sexual risk in an HIV endemic community in Uganda. *AIDS and Behavior*, 15(7):1512–1519, 2011.

[89] Frances C. Moore and Delavane B. Diaz. Temperature impacts on economic growth warrant stringent mitigation policy. *Nature Climate Change*, 5(2):127–131, 2015.

[90] Erin A Mordecai, Krijn P Paaijmans, Leah R Johnson, Christian Balzer, Tal Ben-Horin, Emily Moor, Amy McNally, Samraat Pawar, Sadie J Ryan, Thomas C Smith, et al. Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology letters*, 16(1):22–30, 2013.

[91] Valerie Mueller, Clark Gray, and Katrina Kosec. Heat stress increases long-term human migration in rural Pakistan. *Nature climate change*, 4(3):182–185, 2014.

[92] George C Myers, Robert McGinnis, and George Masnick. The duration of residence approach to a dynamic stochastic model of internal migration: a test of the axiom of cumulative inertia. *Eugenics quarterly*, 14(2):121–126, 1967.

[93] Linda Nash. From safety to risk: The Cold War contexts of American environmental policy. *Journal of Policy History*, 29(01):1–33, 2017.

[94] Calistus N Ngonghala, Giulio A De Leo, Mercedes M Pascual, Donald C Keenan, Andrew P Dobson, and Matthew H Bonds. General ecological models for human subsistence, health and poverty. *Nature Ecology & Evolution*, 1(8):1153, 2017.

[95] Calistus N Ngonghala, Mateusz M Pluciński, Megan B Murray, Paul E Farmer, Christopher B Barrett, Donald C Keenan, and Matthew H Bonds. Poverty, disease, and the ecology of complex systems. *PLoS biology*, 12(4):e1001827, 2014.

[96] Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, pages 1417–1426, 1981.

[97] William Nordhaus and Paul Sztorc. Dice 2013r: Introduction and user's manual. *retrieved November*, 2013.

[98] William D Nordhaus. Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, page 201609244, 2017.

[99] Department of Civil and Environmental Engineering/Princeton University. Global meteorological forcing dataset for land surface modeling. research data archive at the national center for atmospheric research, computational and information systems laboratory, 2006.

[100] Joint United Nations Programme on HIV/AIDS et al. Global aids update 2016. *Geneva: UNAIDS*, 2016.

[101] Brian C O'Neill and Michael Oppenheimer. Dangerous climate impacts and the Kyoto Protocol. *Science*, 296(5575):1971–1972, 2002.

[102] Michael Oppenheimer and Annie Petsonk. Article 2 of the UNFCCC: Historical origins, recent interpretations. *Climatic change*, 73(3):195–226, 2005.

[103] Stephen Pacala and Robert Socolow. Stabilization wedges: solving the climate problem for the next 50 years with current technologies. *science*, 305(5686):968–972, 2004.

[104] Sophie JS Pascoe, Lisa F Langhaug, Webster Mavhu, James Hargreaves, Shabbar Jaffar, Richard Hayes, and Frances M Cowan. Poverty, food insufficiency and hiv infection and sexual behaviour among young rural Zimbabwean women. *PLoS One*, 10(1):e0115290, 2015.

[105] M Pascual, B Cazelles, M.J Bouma, L.F Chaves, and K Koelle. Shifting patterns: malaria dynamics and rainfall variability in an African highland. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1631), 2008.

[106] M. Pascual, L. F. Chaves, B. Cash, X. Rodó, and Md Yunus. Predicting endemic cholera: The role of climate variability and disease dynamics. *Climate Research*, 36(2):131–140, 2008.

[107] Judea Pearl. *Causality*. Cambridge university press, 2009.

[108] Thomas C Peterson, Martin P Hoerling, Peter A Stott, and Stephanie C Herring. Explaining extreme events of 2012 from a climate perspective. *Bulletin of the American Meteorological Society*, 94(9):S1–S74, 2013.

[109] Thomas C Peterson, Peter A Stott, and Stephanie Herring. Explaining extreme events of 2011 from a climate perspective. *Bulletin of the American Meteorological Society*, 93(7):1041–1067, 2012.

[110] June J Pilcher, Eric Nadler, and Caroline Busch. Effects of hot and cold temperature exposure on performance: a meta-analytic review. *Ergonomics*, 45(10):682–698, 2002.

[111] R. S. Pindyck. The use and misuse of models for climate policy. *Review of Environmental Economics and Policy*, 11(1):100–114, 2017.

[112] Virginia E Pitzer, Ccile Viboud, Wladimir J Alonso, Tanya Wilcox, C Jessica Metcalf, Claudia A Steiner, Amber K Haynes, and Bryan T Grenfell. Environmental drivers of the spatiotemporal dynamics of respiratory syncytial virus in the United States. *PLoS pathogens*, 11(1):e1004591, 1 2015.

[113] Marc Poumadère, Claire Mays, Sophie Le Mer, and Russell Blong. The 2003 heat wave in France: Dangerous climate change here and now. *Risk Analysis*, 25(6):1483–1494, 2005.

[114] Jerry D Ramsey. Task performance in heat: a review. *Ergonomics*, 38(1):154–165, 1995.

[115] Samuel Randalls. History of the 2c climate target. *Wiley Interdisciplinary Reviews: Climate Change*, 1(4):598–605, 2010.

[116] Jonathan Robinson and Ethan Yeh. Transactional sex as a response to risk in western kenya. *American Economic Journal: Applied Economics*, 3(1):35–64, 2011.

[117] Xavier Rodo, Mercedes Pascual, George Fuchs, and A S G Faruque. ENSO and cholera: a nonstationary link related to climate change? *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12901–6, 10 2002.

[118] G. H. Roe and M. B. Baker. Why Is climate sensitivity so unpredictable? *Science*, 318(5850):629–632, 2007.

[119] P Saddier, D Floret, H A Guess, F Dürr, J C Peyrieux, D J Weber, and S A Plotkin. Cost of varicella in France: a study in day care centers. *The Journal of infectious diseases*, 178 (Suppl:58–63, 1998.

[120] Wolfram Schlenker and David B Lobell. Robust negative impacts of climate change on african agriculture. *Environmental Research Letters*, 5(1):014010, 2010.

[121] O Seppanen, William J. Fisk, and Q.H. Lei. Effect of temperature on task performance in office environment. *Lawrence Berkeley National Laboratory*, 7 2006.

[122] O Seppanen, WJ Fisk, and D Faulkner. Control of temperature for health and productivity in offices. *Lawrence Berkeley National Laboratory*, 2004.

[123] J Shaman and M Kohn. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3243–3248, 2009.

[124] Jeffrey Shaman, Virginia E Pitzer, Ccile Viboud, Bryan T Grenfell, and Marc Lipsitch. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS biology*, 8(2):e1000316, 2 2010.

[125] Kate Shannon, Thomas Kerr, MJ Milloy, Aranka Anema, Ruth Zhang, Julio SG Montaner, and Evan Wood. Severe food insecurity is associated with elevated unprotected sex among hiv-seropositive injection drug users independent of haart use. *AIDS (London, England)*, 25(16):2037, 2011.

[126] Jeffrey Shrader. Expectations and adaptation to environmental risks. Technical report, Working paper, 2016.

[127] Robert Socolow. Climate change and destiny studies: Creating our near and far futures. *Bulletin of the Atomic Scientists*, 71(6):18–28, 2015.

[128] E Somanathan. *The Impact of Temperature on Productivity and Labor Supply: Evidence from Indian Manufacturing*. PhD thesis, Delhi School of Economics, 2015.

[129] Thomas Stocker. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2014.

[130] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500, 2012.

[131] Nassim Nicholas Taleb. Black swans and the domains of statistics. *The American Statistician*, 61(3):198–200, 2007.

[132] Anna Talman, Susan Bolton, and Judd L Walson. Interactions between hiv/aids and the environment: toward a syndemic framework. *American journal of public health*, 103(2):253–261, 2013.

[133] James Tamerius, Cecile Viboud, Jeffrey Shaman, and Gerardo Chowell. Impact of school cycles and environmental forcing on the timing of pandemic influenza activity in Mexican states, may-december 2009. *PLoS Computational Biology*, 11(8), 2015.

[134] D. E. te Beest, M. van Boven, M. Hooiveld, C. van den Dool, and J. Wallinga. Driving factors of influenza transmission in the Netherlands. *American Journal of Epidemiology*, 178:1469–1477, 2013.

[135] Claudia Tebaldi and Reto Knutti. The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 365(1857):2053–2075, 2007.

[136] Alexander C Tsai, Kristin J Hung, and Sheri D Weiser. Is food insecurity associated with hiv risk? cross-sectional evidence from sexually active women in brazil. *PLoS Medicine*, 9(4):e1001203, 2012.

[137] UNEP UNAIDS. Climate change and AIDS: A joint working paper. 2008.

[138] Arely Vergara-Castañeda, Alejandro Escobar-Gutiérrez, Karina Ruiz-Tovar, Julio Sotelo, Graciela Ordoñez, Mayra Yolanda Cruz-Rivera, Salvador Fonseca-Coronado, Armando Martinez-Guarneros, Juan C. Carpio-Pedroza, and Gilberto Vaughan. Epidemiology of varicella in Mexico. *Journal of Clinical Virology*, 55(1):51–57, 9 2012.

[139] Emilia Vynnycky and Richard White. *An introduction to infectious disease modelling*. Oxford University Press, 2010.

[140] Sheri D Weiser, Karen Leiter, David R Bangsberg, Lisa M Butler, Fiona Percy-de Korte, Zakhe Hlanze, Nthabiseng Phaladze, Vincent Iacopino, and Michele Heisler. Food insufficiency is associated with high-risk sexual behavior among women in botswana and swaziland. *PLoS medicine*, 4(10):e260, 2007.

[141] H. M. Yang, M. L. G. Macoris, K. C. Galvani, M. T. M. Andrighetti, and D. M. V. Wanderley. Assessing the effects of temperature on dengue transmission. *Epidemiology and Infection*, 137(08):1179, 8 2009.

[142] Lei Zhao, Xuhui Lee, Ronald B Smith, and Keith Oleson. Strong contributions of local background climate to urban heat islands. *Nature*, 511(7508):216–9, 7 2014.

[143] Joshua Graff Zivin and Matthew J Neidell. Temperature and the allocation of time: Implications for climate change. Technical report, National Bureau of Economic Research, 2010.

[144] Joshua S. Graff Zivin, Solomon Hsiang, and Matthew Neidell. Temperature and human capital in the short- and long-run. Technical report, National Bureau of Economic Research, Cambridge, MA, 5 2015.