

**The Economic Development of Puerto Rico After United  
States Annexation**

by

**Brian C. Marein**

B.S., The Ohio State University, 2013

M.A., University of Colorado Boulder, 2017

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Economics

2021

Committee Members:

Taylor Jaworski, Co-chair

Carol Shiue, Co-chair

Murat Iyigun

Wolfgang Keller

Andy Baker

Marein, Brian C. (Ph.D., Economics)

The Economic Development of Puerto Rico After United States Annexation

Thesis directed by Prof. Taylor Jaworski and Prof. Carol Shiue

This dissertation reexamines the economic development of Puerto Rico after annexation by the United States in 1898. I introduce new evidence in four papers exploring complementary aspects of development: public health and the mortality transition, roads and local development, patterns in regional growth, and changes in adult height. The introductory chapter provides a brief overview of recent scholarship.

The second chapter examines public health during one of the fastest mortality transitions in history and the first outside of Europe and Western offshoots. Local health departments caused most of the reduction in infant, tuberculosis, and maternal mortality from 1923 to 1945 without significantly increasing public expenditures. I present descriptive evidence that more per capita nurses and midwives, but not sanitary inspectors, correspond to larger declines in infant and maternal mortality.

The third chapter assesses the effect of roadbuilding on local economic development and regional inequality by studying the effort to connect all towns with roads. Using newly digitized maps, I show that US investments failed to reduce regional disparities. Early access to roads promoted local economic development and gave rise to path dependence in the location of economic activity, although geographic factors determined the general spatial pattern of development.

The fourth chapter describes spatial patterns in population growth from 1765 to 2010. The spatial distribution of population began to resemble the modern distribution after the turn of the 20th century, when municipal population densities diverged. Municipal population growth was positively correlated with crop production in the preindustrial era and was negatively correlated with agricultural employment from 1899 to 1970. Urbanization commenced around 1900, decades before most of the Caribbean and Central America.

The fifth chapter considers the biological standard of living in the early 20th century. Drawing on data from three surveys, I show that male height in Puerto Rico increased at more than twice the average rate for Latin America and the Caribbean between 1890 and 1940. I also show that Puerto Ricans at mid-century were among the tallest Latin Americans outside of Argentina and Uruguay. The evidence supports the conclusion that conditions improved substantially.

## Dedication

To Lilla and Gabriella.

## Acknowledgements

I thank my dissertation committee co-chair Taylor Jaworski for guiding me through all aspects of the economics profession and for repeatedly going out of his way to support me. I thank co-chair Carol Shiue for helping to transform my scattered thoughts into a coherent project and for excellent advice on academic writing. I also benefited from feedback from the other members of my committee: Murat Iyigun, Wolfgang Keller, and Andy Baker. John Devereux gave excellent comments on early drafts of papers. I thank Ann Carlos for cultivating my interest in economic history, getting me started on research, and encouraging me to study topics that matter to me. I am deeply indebted to the Interlibrary Loan Service at the University of Colorado Boulder. I am grateful for Paul Kupiec for checking in and keeping my spirits up throughout graduate school. I am indebted to every educator who has helped me along the road to graduate school, particularly my fourth grade teacher, Bonnie Bernstein, and my high school teacher, Maria (LaGuardia) Krejsa, for inspiring me to study Spanish—a language which enriches my life and without which I never could have studied Puerto Rican history. I thank my parents-in-law, Joe and Ildi, for their unwavering support and enthusiasm. I thank my parents, Mark and Cathy, for everything they have given me—life, love, and education. Finally, I thank my wife, Lilla, and daughter, Gabriella, for their endless love and support and for motivating me throughout graduate school. Lilla also provided invaluable feedback on research.

## Contents

<b>Chapter</b>	
<b>1</b>	<b>1</b>
<b>2</b>	<b>8</b>
2.1	8
2.2	12
2.2.1	12
2.2.2	14
2.3	18
2.4	23
2.5	26
2.5.1	26
2.5.2	28
2.5.3	34
2.6	37
2.7	40
<b>3</b>	<b>43</b>
3.1	43
3.2	46
3.3	49

3.4	Roads and Regional Inequality . . . . .	53
3.5	The Effect of Roads on Local Development . . . . .	58
3.5.1	Grid Data . . . . .	58
3.5.2	Municipal Data . . . . .	64
3.6	Conclusion . . . . .	69
<b>4</b>	<b>Spatial Population Trends and Economic Development in Puerto Rico, 1765-2010</b>	<b>70</b>
4.1	Introduction . . . . .	70
4.2	Regional Development Throughout History . . . . .	73
4.3	Data . . . . .	77
4.4	Municipal Population Density Over Two and Half Centuries . . . . .	80
4.4.1	Persistence in Municipal Population Density . . . . .	80
4.4.2	Variation in Municipal Population Density . . . . .	84
4.4.3	Correlates of Municipal Population Growth . . . . .	87
4.5	Urbanization in Comparative Perspective . . . . .	95
4.6	Conclusion . . . . .	99
<b>5</b>	<b>Economic Development in Puerto Rico After US Annexation: Anthropometric Evidence</b>	<b>101</b>
5.1	Introduction . . . . .	101
5.2	Puerto Rico in the Early 20th Century . . . . .	104
5.3	Data . . . . .	107
5.3.1	Puerto Rico Heart Health Program (PRHHP) . . . . .	108
5.3.2	Puerto Rican Elderly: Health Conditions (PREHCO) Project . . . . .	109
5.3.3	Thieme (1959) . . . . .	110
5.4	Comparisons by Birth Year . . . . .	111
5.4.1	Analysis . . . . .	111
5.4.2	Potential Biases . . . . .	116
5.5	Comparisons by Age . . . . .	119

5.6	Cross-Country Comparison . . . . .	123
5.7	Discussion and Conclusion . . . . .	126
<b>Bibliography</b>		<b>128</b>
<b>Appendix</b>		
<b>A</b>	<b>Supplemental Material for Chapter 2</b>	<b>146</b>
A.1	Robustness Checks . . . . .	146
A.1.1	Random Placebo Test . . . . .	146
A.1.2	Underregistration of Births . . . . .	148
A.1.3	Changes in Fertility . . . . .	150
A.1.4	PHU Dates of Inauguration . . . . .	152
A.1.5	Municipality-Specific Time Trends . . . . .	157
A.1.6	Census Controls . . . . .	157
A.1.7	Controlling for a Public Water System . . . . .	162
A.1.8	Unweighted Observations . . . . .	162
A.1.9	Spillovers between Municipalities . . . . .	163
A.1.10	Spatial Autocorrelation . . . . .	164
A.1.11	Static Difference-in-Differences . . . . .	166
A.1.12	Infant Mortality, Maternal Mortality, and Midwives (Alternative Specification)	168
A.2	Heterogeneity by Socioeconomic Characteristics . . . . .	170
A.3	Additional Tables . . . . .	173
<b>B</b>	<b>Supplemental Material for Chapter 3</b>	<b>175</b>
B.1	Data . . . . .	175
B.1.1	Dependent Variables . . . . .	175
B.1.2	Development Controls . . . . .	176

B.1.3	Transportation Data . . . . .	177
B.1.4	Geographic Controls . . . . .	178
B.2	Additional Tables . . . . .	180
B.3	Additional Figures . . . . .	186
B.4	Spatial Autocorrelation . . . . .	187
<b>C</b>	<b>Supplemental Material for Chapter 4</b>	<b>189</b>
C.1	Data . . . . .	189
C.1.1	Dependent Variables . . . . .	189
C.1.2	Development Variables . . . . .	190
C.1.3	Geographic Variables . . . . .	190
C.2	Standardization of Municipality Boundaries . . . . .	191
C.3	Correlates of Net Migration . . . . .	195
C.4	Metropolitan Areas in Central America and the Caribbean in 1950 . . . . .	198
C.5	Additional Figures . . . . .	199
C.6	Additional Tables . . . . .	201
<b>D</b>	<b>Supplemental Material for Chapter 5</b>	<b>204</b>
D.1	Additional Tables . . . . .	204

## Tables

### Table

2.1	Changes in Mortality by Personnel Per Capita . . . . .	35
2.2	Expenditures on PHUs by the Insular Department of Health, 1931-32 to 1938-39 . . . . .	39
2.3	Cost-benefit Analysis, 1931/32-1938/39 . . . . .	40
3.1	Descriptive Statistics: Grid Cells . . . . .	50
3.2	Descriptive Statistics: Municipalities . . . . .	51
3.3	Population Density in Year $t$ Relative to Distance to Roads in 1899 . . . . .	54
3.4	Modern Development Outcomes and Access to the Transport Network in 1899, 1920, and 1940 . . . . .	56
3.5	Persistent Effects of Historical Roads on Economic Outcomes—OLS Results . . . . .	59
3.6	Persistent Effects of Historical Roads on Economic Outcomes—2SLS Results . . . . .	62
3.7	Persistent Effects of Historical Market Access on Economic Outcomes—OLS Results . . . . .	65
3.8	Persistent Effects of Historical Market Access on Economic Outcomes—2SLS Results (Mean Travel Time) . . . . .	67
4.1	Descriptive Statistics . . . . .	79
4.2	Variation in Population Density . . . . .	85
4.3	Correlates of Population Growth . . . . .	90
5.1	Mean Height of Puerto Rican Men by Birth Year . . . . .	109
5.2	Changes in Height Across Birth Cohorts in the PRHHP and PREHCO . . . . .	113

5.3	Mean Height of Puerto Rican Men by Age . . . . .	121
5.4	Mean Male Height in Latin America, Spain, and the United States . . . . .	125
A.1	Robustness Checks: PHU Dates of Inauguration (Infant Mortality) . . . . .	153
A.2	Robustness Checks: PHU Dates of Inauguration (Tuberculosis Mortality) . . . . .	154
A.3	Robustness Checks: PHU Dates of Inauguration (Overall Mortality) . . . . .	155
A.4	Robustness Checks: PHU Dates of Inauguration (Stillbirths) . . . . .	156
A.5	Robustness Checks: Alternative Specifications (Infant Mortality) . . . . .	158
A.6	Robustness Checks: Alternative Specifications (Tuberculosis Mortality) . . . . .	159
A.7	Robustness Checks: Alternative Specifications (Overall Mortality) . . . . .	160
A.8	Robustness Checks: Alternative Specifications (Stillbirths) . . . . .	161
A.9	The Effect of PHUs on Mortality Rates Controlling for Neighboring Units . . . . .	163
A.10	The Effect of Public Health Units on Mortality Rates by Cause of Death (Standard Errors Adjusted for Spatial Autocorrelation) . . . . .	165
A.11	The Effect of Public Health Units on Mortality Rates (Difference-in-Differences Specifications) . . . . .	167
A.12	Treatment Heterogeneity of Public Health Units by Municipal Characteristics . . . . .	172
A.13	Selected Activities of Public Health Units in Fiscal Year 1938-39 . . . . .	173
A.14	Leading Causes of Death, 1931-32 . . . . .	174
A.15	Mortality Rates per 100,000 Inhabitants for the Principal Infectious and Parasitic Diseases, 1931-1949 . . . . .	174
B.1	Persistent Effects of Historical Market Access on Economic Outcomes: 1910 Agricultural Controls . . . . .	181
B.2	Persistent Effects of Historical Market Access on Economic Outcomes: FAO Land Suitability . . . . .	182
B.3	Persistent Effects of Historical Market Access on Economic Outcomes: 1899 Development Controls . . . . .	183

B.4	Persistent Effects of Historical Market Access on Economic Outcomes: Different Values of $\theta$ . . . . .	184
B.5	Persistent Effects of Historical Market Access on Economic Outcomes—2SLS Results (Maximum Travel Time) . . . . .	185
C.1	Example of Aggregation of City Populations: Barceloneta, Florida, and Manatí . . .	193
C.2	Municipality Standardization Decisions . . . . .	194
C.3	Municipality Standardization Decisions (cont.) . . . . .	195
C.4	Correlates of Net Migration . . . . .	197
C.5	Correlates of Population Growth (Unweighted) . . . . .	202
C.6	Correlates of Population Growth (Alternative Geographic Variables) . . . . .	203
D.1	Changes in Height Across Birth Cohorts in the PRHHP and PREHCO (Quantile Regressions) . . . . .	205
D.2	Changes in Predicted Height Across Birth Cohorts in the 1970 US and Puerto Rican Censuses . . . . .	206
D.3	Average (Imputed) Years of Education of Puerto Rican Men by Birth Year . . . . .	207

## Figures

### Figure

2.1	The Largest Improvement in Life Expectancy Over 30 Years by Country . . . . .	9
2.2	Deaths Per 1,000 Inhabitants in Puerto Rico, 1888-1960 (3-Year Moving Average) . . . . .	13
2.3	The Roll-out of Public Health Units . . . . .	19
2.4	Mortality Rates, 1923-1945 . . . . .	21
2.5	The Effect of PHUs on Mortality: Baseline Specification . . . . .	27
2.6	The Effect of PHUs on Mortality: Dynamic Specification . . . . .	30
2.7	The Effect of PHUs on Mortality: de Chaisemartin and D'Haultfoeuille Estimator . . . . .	32
2.8	Resources of the Department of Health, 1928-29 to 1942-43 (Thousands of dollars) . . . . .	38
3.1	Transport Network in 1899 . . . . .	57
3.2	Population Density by <i>Barrio</i> in 1899 . . . . .	57
3.3	Population Density by <i>Barrio</i> in 2010 . . . . .	57
3.4	Market Access and Population Density, 1846-2010 . . . . .	68
4.1	Shares of Total Export Value by Product, 1865-1929 . . . . .	74
4.2	Labor Force by Industry (in thousands), 1899-2000 . . . . .	76
4.3	Map of Standardized Municipalities ( <i>Municipios</i> ) in 1827 . . . . .	79
4.4	Municipal Population Density (per sq. km.), 1827-2010 (Selected Years) . . . . .	81
4.5	Correlation Between Historical Municipal Population Density and Municipal Population Density in 2010 . . . . .	83

4.6	Development Variables and Population Growth by Census Year . . . . .	93
4.7	Population Growth and Contemporaneous Municipal Literacy Rates . . . . .	94
4.8	An Urbanization Index for Central America and the Caribbean . . . . .	97
5.1	Mean Height of Puerto Rican Men by Birth Year and Age . . . . .	122
A.1	The Distributions of Coefficients From Random Placebo Tests . . . . .	147
A.2	The Effect of Public Health Units on the Ratio of Infant Deaths to Population . . .	149
A.3	The Effect of Public Health Units on Fertility . . . . .	151
A.4	Treatment Heterogeneity by Midwives Per Capita (Infant and Maternal Mortality) .	169
B.1	Transport Network in 1899 and Straight Line Instruments . . . . .	186
B.2	Transport Network in 1920 . . . . .	186
B.3	Transport Network in 1940 . . . . .	187
C.1	Development Variables and Population Growth (Unweighted) . . . . .	200

## Chapter 1

### Introduction

The economic history of Puerto Rico tends to be neglected by scholars outside of Puerto Rican studies. On the one hand, Puerto Rico is usually excluded from discussions of Latin America because its special relationship with the United States has shaped a very different pattern of development (e.g. Bulmer-Thomas, 2003). On the other hand, Puerto Rico is left out of, or hidden, from the story of the United States because it is an unincorporated territory, not a state (Immerwahr, 2019). Puerto Rican migrants to the mainland have been treated as foreigners by their fellow citizens (Fernández, 2010), and nearly half of American adults were unaware that Puerto Ricans were US citizens as recently as 2017 (Morning Consult, 2017). Puerto Rico does not fall neatly into Anglo- or Latin America and is, therefore, often ignored.

Yet, Puerto Rico has much to offer, particularly to economic historians and development economists, *because* it is situated between two worlds. Puerto Rico is poorer than all US states but richer than most of Latin America (Devereux, 2019). How did this happen? As late as the early 19th century Puerto Rico relied mostly on subsistence agriculture, lagging behind most of the West Indies (Mintz, 1974). Even after the rise of commercial agriculture, Puerto Rico remained one of the less developed areas of Latin America, as gross domestic product per capita declined over the late 19th century (Santamaría García, 2012). Following annexation by the United States in 1898, the island emerged as one of the leading sugar producers in the world, then rapidly industrialized and ranked among the fastest growing economies after World War II (Baumol and Wolff, 1996).

Of central importance to the economic history of Puerto Rico is annexation by the United

States in 1898, ending four centuries of Spanish rule. Scholars have reached a broad range of conclusions about US annexation, from “extreme denunciations of US imperialism and its purported ruthless exploitation and impoverishment of the island...” to “unabashed praises of the US invasion for having saved Puerto Rico from the tyranny of Spanish colonialism, for building a modern economic infrastructure, and for eventually ushering in an era of democracy, relative prosperity, and social mobility...” (Ayala and Bergad, 2002, p. 65).

A key challenge to resolving the debate on economic conditions in early 20th century Puerto Rico is a lack of quantitative evidence. Thankfully, renewed interest in Puerto Rico has produced new evidence to establish basic facts about this era. Devereux (2019) introduces a GDP index for 1900 to 1940 and argues that income per capita grew at impressive rates during direct American rule, in contrast to the prevailing view in the literature. Ayala and Bergad (2020) convincingly show with a large dataset of land tax records that the narrative—found in nearly every major work of Puerto Rican history—that large-scale, absentee-owned US sugar-manufacturing corporations displaced a small landholding peasantry has no basis in fact; instead, economic growth and free trade with the United States resulted in the expansion of the number of farms and farm owners.

I follow these works in reexamining the economic development of Puerto Rico after US annexation, introducing quantitative evidence to compare conditions across Spanish and US rule. Each chapter explores a complementary aspect of development: public health and the mortality transition, colonial roads and local development, changes in adult height, and patterns in regional growth. I contribute rich new data on public health and transport infrastructure in the early 20th century, assemble several underused sources of anthropometric data, and standardize more than two centuries of municipal-level census data. I argue that living conditions improved substantially between US annexation in 1898 and US (and Puerto Rican) entry into World War II in 1941 and that investments in public goods and services—namely, public health and transport infrastructure—were substantial before the postwar industrialization and limited home rule.

In Chapter 2, I examine public health and changes in mortality from 1923 to 1945. Between 1930 and 1960, life expectancy increased by 28.8 years—a rate unprecedented, and since rarely

exceeded, in world history—and achieved parity with the United States. Puerto Rico completed the first mortality transition outside of Europe and Western offshoots. Because the decrease in mortality accelerated while growth of GDP per capita ranked among the fastest rates in the world, there is disagreement about whether public health mattered (Davis, 1953; Molina and Noam, 1964; Stolnitz, 1953; Vázquez Calzada, 1984). While not ruling out the importance of economic growth, this chapter highlights the role played by public health during the early years of the mortality transition and confirms that sustained reductions in mortality began during the Great Depression, prior to the postwar expansion. My results also contribute to a debate regarding the efficacy of anti-tuberculosis (TB) efforts before the advent of antibiotics, which can inform modern public policy to address the crisis of drug-resistant TB in developing countries (Hansen et al., 2020).

Specifically, I use municipal-level data in an event study framework to show that public health units (PHUs, or county health departments) caused most of the reduction in infant, TB, and maternal mortality and one-fifth of the decline in general mortality from 1923 to 1945. The evidence also suggests that PHUs reduced stillbirths but, consistent with the historical record, had no effect on malaria mortality. PHUs carried out a wide range of activities, making it difficult to know which efforts were responsible for reducing mortality. I present descriptive evidence based on cross-sectional data from 1935 that municipalities with more per capita nurses and assistant midwives, but not sanitary inspectors, experienced larger declines in infant and maternal mortality.

Remarkably, PHUs opened during the Great Depression, when external aid was steadily diminishing, and without significantly increasing public expenditures (Payne, 1942). The Department of Health continued the rollout of PHUs despite a shrinking budget, including a 29 percent cut from 1931-32 to 1933-34, by simplifying its organization and cutting waste. Notably, despite significant contributions from the federal government and the Rockefeller Foundation, PHUs were administered primarily by local doctors and nurses and funded mostly by domestic taxes.

In Chapter 3, I examine the effect of road-building on local economic development and regional inequality by studying the effort in the early 20th century to connect all cities and towns with roads. Good roads were deemed “fundamental to the well-being and progress” of Puerto

Rico, providing access to schools and facilitating commerce (Carroll, 1899, p. 40). From the outset, officials emphasized the urgency of providing “good roads from seaboard towns into the interior and of substantial trails branching out into the less populous communities” to develop areas lacking transport facilities (Governor of Porto Rico, 1901, p. 427). The construction of roads to benefit the local economy conflicts with claims that public investments were intended to narrowly benefit American sugar corporations (e.g., Picó, 1983) or to project military power (Jackson, 2020).

Using newly digitized historical maps of the transport network and district-level population data, as well as several measures of modern development, I consider how historical investments in roads shaped the spatial distribution of economic activity in the short and long run. In 1899, the year of the first US census, roads connected the largest cities and towns, but proximity to roads does not strongly predict population density away from the network nodes. After 1899, the population increasingly clustered near old roads built under Spanish rule—and modernized under early US rule—despite the rapid construction of roads elsewhere. Areas near old Spanish roads became, and remain, richer and more densely populated than other areas, revealing that investments made under early US rule failed to reduce spatial inequality.

Additionally, I employ an instrumental variables strategy to identify the effect of historical roads on local development, exploiting plausibly exogenous variation in access to roads based on the locations of historical department capitals. My results indicate that early investments in roads promoted local economic development and that their effect increased during the rise of commercial agriculture and external trade in the early 20th century, giving rise to path dependence in the location of economic activity. However, several other variables (i.e. historical population density, distance to San Juan, elevation, and suitability for sugarcane cultivation) are stronger predictors of modern development than distance to historical roads; roads influenced the precise location of economic activity, but geography shaped the general pattern of development. My findings suggest that temporary advantages can have small but persistent effects on city size, so investments in infrastructure intended to disperse development may be ineffective.

In Chapter 4, I describe spatial patterns in population growth in Puerto Rico from 1765 to

2010 using data for consistently defined municipalities. First, I show that population density throughout the 18th and 19th centuries is weakly correlated with modern population density. After 1899, the distribution of population across municipalities quickly approached the modern distribution, as the greatest change in relative population density occurred between 1899 and 1910. These results indicate that the turn of the century—not just the postwar period, as suggested in previous research—was a turning point for economic development. Second, I find that variation in municipal population density was high in the 18th century, fell throughout the 19th century as the population spread across the island, began to increase around the turn of the 20th century as population clustered in and near San Juan, accelerated after 1940, and plateaued around 1980. Third, population growth was positively correlated with traditional crop production—coffee in the late 19th century, and sugarcane and tobacco in the early 20th century—in the preindustrial era and was negatively correlated with agricultural employment from 1899 to 1970.

One of the major implications of these results is that urbanization began around 1900—decades earlier than generally believed—when variation in municipal population density began to rise and population growth was higher in areas less reliant on agriculture. Past research tends to treat post-WWII industrialization as a sharp break from a stagnant past (e.g., Baer, 1959). Yet, the spatial trends exhibited during the postwar “miracle” were a mere continuation of trends originating at the turn of the century, indicating that Puerto Rico’s impressive socioeconomic transformation in the 20th century began, contrary to most accounts, before World War II. Moreover, by constructing a time series for an urbanization index using standardized definitions of metropolitan areas proposed by International Urban Research (1959), I show that most of Central America and the Caribbean did not urbanize until after World War II. The conclusion that urbanization in Puerto Rico began around 1900, and that urbanization in many small countries in the Circum-Caribbean lagged by several decades, motivates additional research.

In Chapter 5, I draw on anthropometric evidence from three surveys of Puerto Rican men to comment on economic development. Increases in height have been found to be associated with rising living standards. Height provides additional information about development because it

captures how living conditions changed throughout the entire population; economic growth can fail to translate to improvements in living conditions for vast sections of the population if income inequality rises as well. Height is positively correlated with the economic standard of living, broadly defined, and may serve as a proxy for the “biological standard of living” when more conventional indicators are unavailable (Fogel, 1994; Komlos and Baten, 1998; Steckel, 1995, 2009). Height is heavily influenced by nutrition and disease during childhood.

In this chapter, I show that the height of Puerto Rican men increased significantly in the early 20th century: men born in 1940 measure about 4.2 cm. taller than those born in 1890. Additionally, I show that Puerto Ricans at mid-century were among the tallest Latin Americans outside of Argentina and Uruguay. My results, therefore, complement the GDP index for 1900 to 1940 introduced by Devereux (2019). This chapter also complements Chapter 2 of this dissertation, since an important source of increasing height historically was the improving disease environment as reflected by the fall in infant mortality (Hatton, 2014).

Lacking an adequate control group and sufficient observations predating US rule, I do not argue that US annexation is responsible for changes across birth cohorts; after all, the 20th century witnessed improvements in living standards and stature across the West (NCD Risk Factor Collaboration (NCD-RisC), 2016). Rather, I more narrowly contend that widespread beliefs about impoverishment caused by US policy, or even stagnant economic conditions in the immediate aftermath of annexation, are incorrect. Such claims have been shaped in large part by failing to consider the realities of Spanish colonialism and by observing that living standards in Puerto Rico were much lower than those of the United States and Western Europe without properly comparing Puerto Rico’s present conditions with the conditions that prevailed at the commencement of US rule (Ayala and Bergad, 2002; Chamber of Commerce of Puerto Rico, 1945). My findings are supportive of a marked improvement in the biological standard of living and are not attributable to selection into the sample or shrinking in old age.

Taken together, these chapters provide a new perspective on Puerto Rican development in the aftermath of US annexation. Much work is left to be done. Understanding the economic

history of Puerto Rico since 1898 is especially important given renewed debate on Puerto Rico's unresolved political status and the nature of its relationship with the United States in response to the Puerto Rican government-debt crisis and the economic challenges following Hurricane Maria. This dissertation is a step in that direction.

## Chapter 2

### Public Health in the First Mortality Transition in the Tropics

#### 2.1 Introduction

The decline in mortality rates over the past two centuries ranks among the most significant achievements in history, greatly contributing to human welfare. Life expectancy more than doubled, and children die at rates far below historical averages (Lee, 2003). There is debate in the social sciences regarding the relative contributions of public health and economic growth to this process prior to modern medicine (Anderson et al., forthcoming; Costa, 2007; Cutler and Miller, 2005; Fogel, 2004; McKeown, 1979; Preston, 1980), and the conclusions from this research have important implications for parts of the developing world where contagious and infectious diseases are more prevalent and infant mortality remains high. In particular, such research may provide insight for combating tuberculosis (TB) without antibiotics, since TB remains a leading cause of death worldwide, and drug-resistant forms of the disease exist.

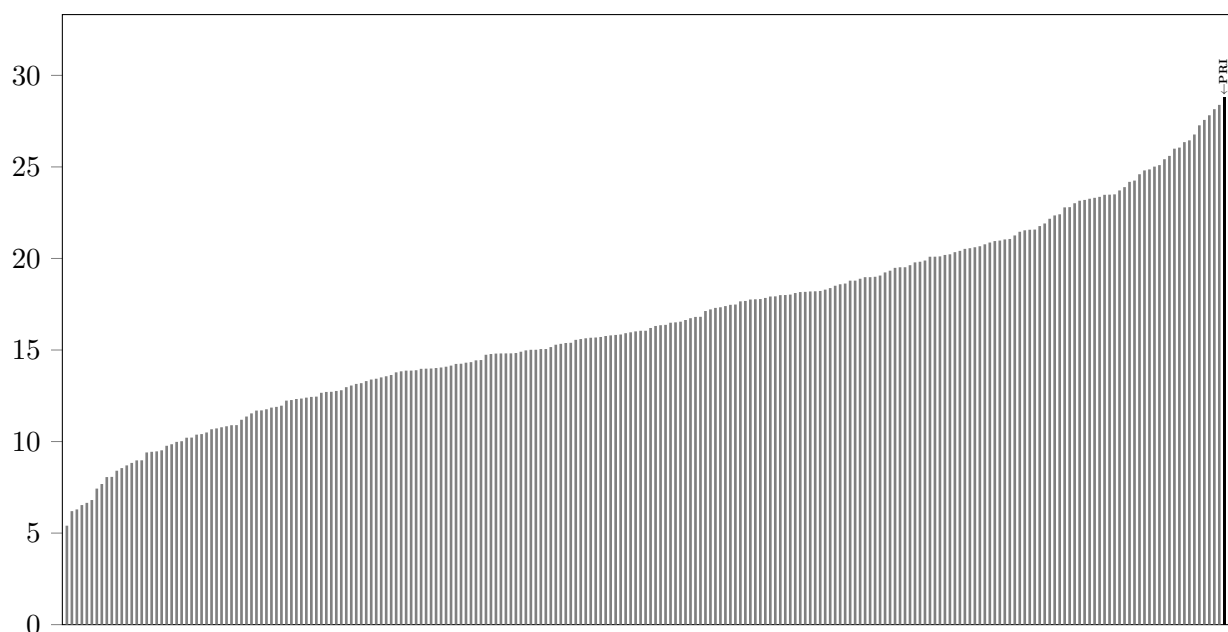
In this paper, I assess the role of public health in the mortality transition in Puerto Rico. Despite a GDP per capita less than one-fifth of the US level and a predominately agrarian society, Puerto Rico completed the first mortality transition outside of Europe and Western offshoots, doing so with public health administered primarily by local doctors and nurses and funded mostly by domestic taxes.<sup>1</sup> The mortality transition in Puerto Rico ranks among the most successful: between 1930 and 1960, life expectancy increased by 28.8 years—a rate unprecedented, and since

---

<sup>1</sup> Sixty percent of the labor force was employed in agriculture and 78 percent of the population resided in rural areas in 1920 (Vázquez Calzada, 1978). Urban areas were defined by the Census as cities, towns, and villages with 2,500 or more inhabitants. Comparisons of GDP per capita are based on data from Devereux (2019).

rarely exceeded, in world history—and achieved parity with the United States (see Fig. 2.1). Puerto Rico overcame a high disease burden typical of the tropics, where health outcomes are worse even after controlling for per capita income (Sachs, 2001), and one of the highest rates of TB mortality in the world (Yelton, 1946), leading a commissioner of health to declare that “the progress of Puerto Rico in public health has no parallel in any other country” (Bourne and Bourne, 1966, p. 81).

Figure 2.1: The Largest Improvement in Life Expectancy Over 30 Years by Country



Sources: Astorga et al. (2005); Roser et al. (2020); Vázquez Calzada (1978)

Notes: Each bar represents the largest increase in life expectancy from birth in years over any thirty year period from 1543-2019 for which data are available for all countries and territories in Roser et al. (2020). Recoveries from large negative health shocks, such as war, are excluded.

Using newly digitized, municipal-level data from 1923 to 1945 in an event study framework, I estimate the decline in mortality attributable to public health units (PHU), or subdivisions of the Department of Health designed to improve local coordination of public health education, control of communicable diseases, and child hygiene, primarily through preventative care and education. I exploit the rollout of the program to identify the effect of PHUs on mortality, as a unit opened in each municipality between 1926 and 1937. I estimate that PHUs were responsible for most of the decline in infant and TB mortality and about one-fifth of the decline in overall mortality

during the initial phase of the mortality transition. PHUs also reduced stillbirths and maternal mortality but, consistent with the historical record, had no effect on malaria mortality. Using a cross-section of personnel from 1935, I find that more per capita nurses and midwives, but not sanitary inspectors, are associated with steeper declines in infant and maternal mortality, suggesting the importance of, e.g., home visits, prenatal clinics, and occupational licensing. The results are robust to allowing for heterogeneity in treatment effects over time or across municipalities (de Chaisemartin and D’Haultfoeuille, 2020) and to controlling for time-varying demographic characteristics, municipality-specific linear trends, and the existence of a public water system.

Modeled after county health departments in the American South, PHUs were a joint effort by municipalities, the Puerto Rico Department of Health, and the Rockefeller Foundation. As the objectives of the Department of Health expanded beyond sanitary police and the suppression of epidemics, officials recognized the inefficiency of having every activity controlled by a separate arm of the insular government in the capital, San Juan (Pons, 1952). The PHU was implemented to unify control over health work under a full-time medical officer who could more effectively identify and respond to local problems. Puerto Rico prioritized the establishment and expansion of PHUs with the confidence that they, “more than anything else,” could resolve ongoing public health problems (Governor of Puerto Rico, 1932, p. 6). Remarkably, PHUs opened during the Great Depression, when aid from outside sources was steadily diminishing, and without significantly increasing public expenditures (Payne, 1942). The Department of Health continued the rollout of PHUs despite a shrinking budget, including a 29 percent cut from 1931-32 to 1933-34, by simplifying its organization and eliminating unnecessary expenses. A back-of-the-envelope calculation indicates that the benefits of PHUs exceeded costs by a ratio of 13 to 1.

The mortality data introduced in this paper offer a glimpse into historical public health outside of Europe and Western offshoots. Data this granular and consistently reported are rare for a developing country from this time period (Preston, 1980). Even today, roughly 70 percent of the world’s population lives in areas without complete registration of deaths (Mahapatra et al., 2007). As a US territory, Puerto Rico published reliable vital statistics, unlike most countries and

territories in Africa, Asia, and Latin America. After Puerto Rico, practically all countries of Latin America adopted public health units, so these data allow for a window into public health in Latin America more broadly (Bravo, 1958). Puerto Rican infant mortality data from 1927 to 1943 are used by McEniry (2011, 2014), but the present analysis builds on the data with additional outcome variables and years covered and is the first effort to explain changes in mortality. These data complement US county-level mortality data digitized by Hoehn-Velasco (2018) and Bailey et al. (2018).

This paper contributes to an ongoing debate on the contributions of public health and economic growth in the mortality transition prior to modern medicine (Anderson et al., forthcoming; Costa, 2007; Cutler and Miller, 2005; Fogel, 2004; McKeown, 1979). Studying Puerto Rico enriches the discussion by introducing evidence from outside of Europe and Western offshoots. I also bring a debate on Puerto Rico into the wider literature: because the decrease in mortality accelerated while growth of GDP per capita ranked among the fastest rates in the world, there is disagreement about whether public health mattered (Davis, 1953; Molina and Noam, 1964; Stolnitz, 1953; Vázquez Calzada, 1984). While not ruling out the importance of economic growth, this paper highlights the role played by PHUs and confirms that sustained reductions in mortality began during the Great Depression, prior to the postwar expansion.

More narrowly, this paper contributes to a debate regarding the efficacy of anti-TB efforts before the advent of antibiotics. Such research is especially consequential because public health policy is informed by historical solutions to address the crisis of drug-resistant TB (Dheda and Migliori, 2012). Evidence of large reductions in TB caused by PHUs is consistent with Bauernschuster et al. (2019), Hollingsworth (2014), and Hansen et al. (2020) but stands in contrast to Anderson et al. (2019a) and Clay et al. (2020), who conclude that early efforts to combat TB in the United States found limited success. Pinpointing the measures that mattered is difficult because PHUs carried out many activities, but the evidence from this investigation together with Hollingsworth (2014) and Hansen et al. (2020) suggest that testing, contact tracing, isolation, and instructions on how to prevent transmission may have been effective.

This paper follows research on historical public health in the United States (Alsan and Goldin, 2019; Anderson et al., 2019a, 2020a, 2019b; Clay et al., 2019; Ferrie and Troesken, 2008; Kitchens, 2013; Komisarow, 2017; Olmstead and Rhode, 2004), including investments made by the Rockefeller Foundation (Bleakley, 2007). Hoehn-Velasco (2018) and Hoehn-Velasco and Wrigley-Field (2021) are the most closely related research and demonstrate that local health departments led to a modest reduction in mortality in the rural, but not urban, United States in the early 20th century. The conclusions of this paper also complement recent studies on public health efforts led by nurses (e.g., home visits and clinics for new mothers, as in Puerto Rico) (Bhalotra et al., 2017; Bütikofer et al., 2019; Moehling and Thomasson, 2014; Wüst, 2012) and the licensing of midwives (Anderson et al., 2020a; Kotsadam et al., 2017; Lazuka, 2018; Pettersson-Lidbom, 2015) in the United States and Scandinavia in the early 20th century.

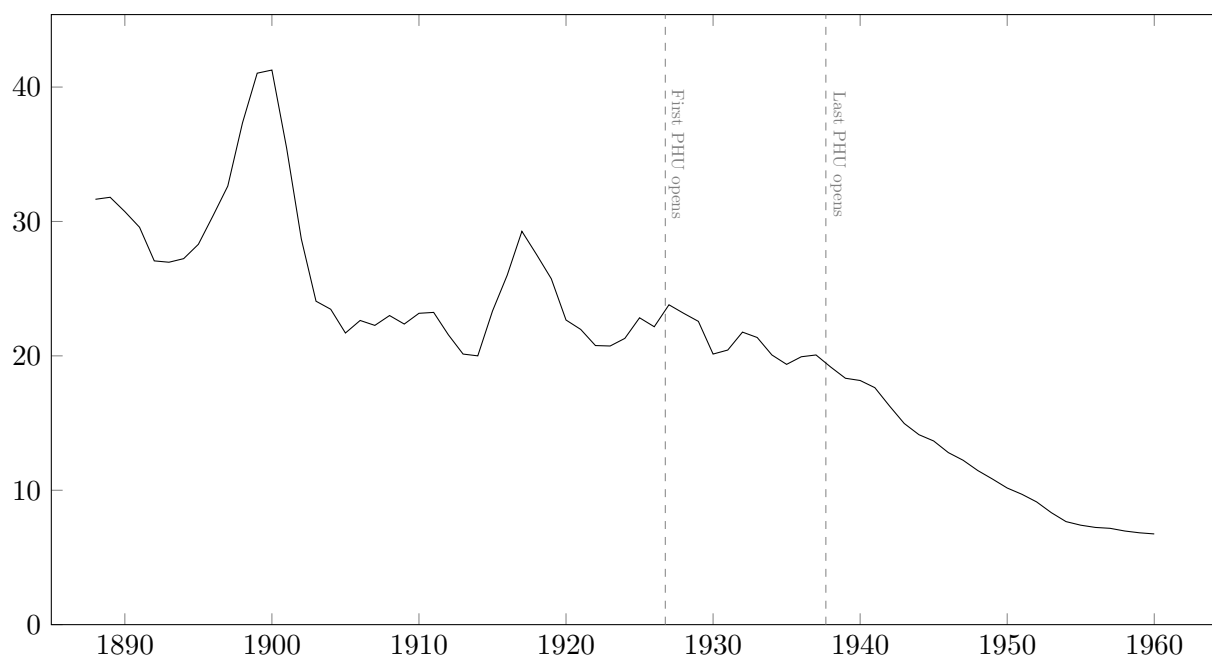
## **2.2 Public Health Units in Puerto Rico**

### **2.2.1 Program Overview**

After a wave of public health investments and a steep drop in mortality following annexation by the United States in 1898, Puerto Rico experienced little change in mortality between 1910 and 1930 (see Fig. 2.2). As the Department of Health expanded the scope of its activities in the early 1920s to address ongoing problems, officials perceived certain defects in the operation of local health work, with each activity controlled by a separate bureau in San Juan and without coordination within municipalities (Pons, 1952). In 1925, the Rockefeller Foundation persuaded Puerto Rico to adopt the system of county health departments found in the United States. The new organizations were to be called public health units and were designed to unify the control of those activities related to the prevention of disease and the conservation of health in a local organization under a trained health officer.

Included in the work of each unit were the study and control of transmissible diseases; health education; medical inspection of schoolchildren; health promotion by special prenatal, infant,

Figure 2.2: Deaths Per 1,000 Inhabitants in Puerto Rico, 1888-1960 (3-Year Moving Average)



Source: Puerto Rico Department of Health

preschool and TB clinics; and the control of general sanitation (Chaves, 1938c). Some activities had previously been performed by multiple agencies operating under central control but lacking local coordination, but the rapid advance of the objectives of the Department of Health motivated the creation of a new institution for carrying out the more complicated procedures of modern health work. For example, prior to the opening of the first PHU, public health nursing was effectively nonexistent, as nurses were only employed in hospitals for bedside care (Chaves, 1938b). Physical examinations of schoolchildren were conducted in few municipalities before 1926 but were universal by 1937 (Arbona, 1937).

In fiscal year 1938-39, the first full year in which all municipalities were served, PHUs gave 25,873 public lectures to 717,057 attendees, made 164,623 nurse home visits, and examined 170,576 schoolchildren, for a population of approximately 1.8 million. About one in four prenatal cases received care in PHUs that year, and children regularly attended clinics for years past infancy (Arbona, 1940). PHU buildings were located in town centers, but their catchment areas included both the urban and rural populations (Ramírez de Arellano, 1981). Nurses made house calls in the

rural hinterland, and X-ray traveling units diagnosed TB cases in zones without a fixed X-ray unit.

PHUs coordinated all the health efforts of the community, including insular and municipal as well as public and private efforts. Perhaps the most important complement to PHUs were milk stations, which commenced operations in 1930 and aimed to educate mothers in infant feeding and care and provided them with milk of good quality at a nominal cost. The Department of Health estimated that mortality among infants attending the stations was 57 percent lower than general infant mortality over the island (Governor of Puerto Rico, 1932). Milk stations were nominally independent of PHUs but in practice were integrated with PHU services like prenatal care and screening for tuberculosis (González, 2015). In fact, milk stations were often housed in the same facility as PHUs: in 1938-39, 65 of the 68 milk stations were in PHUs, but this proportion varied over time.

Initially, municipalities paid a quarter of their PHU budgets, with the Rockefeller Foundation covering another quarter and the Department of Health covering half, and the share borne by municipalities was to grow over time. Other funding sources included the US Public Health Service, the Federal Emergency Relief Administration, and the Puerto Rico Reconstruction Administration. Early PHUs served single municipalities, but many municipalities could not afford their own units. Consequently, the island was divided into districts consisting of one to four municipalities based on population, means of communication, and similarity in health problems (Chaves, 1938a). Each district had one or more health officers, while each municipality had its own nurses and inspectors.

### **2.2.2 PHU Interventions**

The leading causes of death in this era were infectious and parasitic diseases, accounting for approximately 60 percent of all mortality at the beginning of the study period. Mortality in Puerto Rico declined because of a reduction in deaths from infectious and parasitic diseases. Trends in mortality rates from chronic disease and external causes remained flat, while the share of all deaths caused by infectious and parasitic diseases fell to 48 percent in the late 1940s (Vázquez Calzada, 1978).

### **2.2.2.1 Infant and Maternal Mortality**

PHUs provided extensive guidance on child hygiene and nutrition. Clinics instructed mothers how to properly bathe the baby, prepare baby formula, and keep utensils clean (Nochera, 1940). Nurses tracked the health of children under the care of milk stations through daily visits with mothers, advising mothers when to bring the child to a doctor or examining the child at home and seeking medical attention when needed (Garrido Morales, 1937). Milk stations had a duty to encourage breastfeeding at least until a child reached nine months, and breastfeeding is associated with lower rates of mortality from diarrhea and enteritis, pneumonia, and bronchitis in developing countries (World Health Organization, 2000). Children at milk stations were examined by the medical officer of the PHU, who prescribed a milk formula in accordance with the age and health of the child.

Some efforts combated both infant and maternal mortality. PHUs administered an intensive program of midwife supervision launched by the Bureau of Infant Hygiene in 1931 after discovering that all children who died of infantile tetanus had been attended at birth by unlicensed midwives “under conditions where cleanliness and asepsis were entirely absent” (Robert de Romeu, 1938, p. 56). The Health Department created a new professional category, the assistant midwife, to differentiate them from midwives who had nursing degrees (del Río, 1938). Their strategy maintained the accessibility of midwifery services while teaching assistant midwives the rudimentary principles of obstetrics and the care of mother and child at home. The law compelled midwives to attend classes once a week at PHUs until a license was granted. By 1938, licensed midwives attended 85 percent of all registered births. PHUs also distributed sterile umbilical dressings to midwives free of charge and required midwives to purchase antiseptic solutions and germicidal soaps.

Finally, PHUs sought to eliminate hookworm infection, which leads to iron-deficiency anemia and in pregnancy is associated with preterm labor, low neonatal weight, and increased newborn and maternal mortality (Camaschella, 2015; Hotez et al., 2004). Hookworm infection had been

practically universal in rural communities, and expectant mothers were tested and treated at prenatal clinics. Sanitary inspectors from the PHUs showed people how to construct latrines. By 1939, practically every rural home had a latrine, and those unwilling to build a latrine were brought before justices of the peace (Washburn, 1939).

#### **2.2.2.2 Tuberculosis**

Pulmonary tuberculosis, a disease caused by bacteria that spread from person to person through the air, was the leading cause of death among adults. PHUs provided education in the modes of spread of the disease and the essential features of prevention through clinics, an extensive house-to-house campaign, and the distribution of pertinent literature (Garrido Morales, 1935). Additionally, the Department of Health operated a closed system of treatment “from which the patient [could not] escape...” (Coryllos, 1937, p. 59). PHUs served as case-finding agencies, screening schoolchildren and attendees at their clinics and conducting house-to-house searches to direct those in contact with a TB patient to testing. From the PHU, patients were sent to an anti-TB center to confirm the diagnosis and subsequently relocated to the nearest TB hospital. Hospitalization served to isolate infected persons to control disease transmission, and the island-wide anti-TB campaign launched in 1934 tripled the number of hospital beds available for TB patients (Rodríguez Pastor and Janer, 1953). Patients with active TB could undergo surgical procedures to collapse the lung and promote cavity closure at the hospital in Río Piedras. Cases that became negative were discharged, referred to their respective anti-TB center, and sent back to their hometowns to remain “under continuous supervision” by their PHU (Coryllos, 1937, p. 59).

#### **2.2.2.3 Malaria**

PHUs provided anti-malarial drugs to those attending their clinics, participated in the destruction of larvae and other mosquito-control work, and employed special malaria inspectors to take surveys, draw blood samples, and treat cases in homes in areas where the disease was most prevalent (Malaret, 1937). However, because of a lack of funds and trained field force, such

activities by and large represented a continuation of the work that had started years ago, “hardly affecting the seriousness of the malaria conditions in the country at large” (Arbona, 1942, p. 257). The eradication of malaria did not meaningfully begin until the 1940s, when the federal government intervened to fight the disease near military installations (Miranda Franco and Casta Vélez, 1997).

#### **2.2.2.4 Mortality from All Causes**

Much of the decline in mortality from all causes can be attributed to decreases in infant, maternal, and TB mortality. Several other activities could have lowered mortality through the direct reduction of disease transmission or by improving the ability to overcome disease. Child hygiene clinics and the school hygiene program offered advice to children regarding a properly balanced diet, exercise, dental care, and hygiene (Nochera, 1940). The clinics also immunized children against smallpox, diphtheria, and typhoid. Teachers were encouraged to report any case of possible communicable disease among their students to the PHU (Arbona, 1937). The departments of health and education made a concerted effort to treat all children with hookworm infection in municipalities where PHUs had been established after examining the feces of schoolchildren (Chaves, 1937b). PHUs instructed students to wear shoes and use latrines to avoid contracting the infection and contaminating the soil. Inspectors enforced sanitary regulations for the building of houses, pure water supply, and the latrine system and worked to root out the adulteration of milk, food, and drugs (Robert de Romeu, 1937).

#### **2.2.2.5 Stillbirths**

Syphilis tests were required in all prenatal clinics, with free treatment provided for all positive cases (Quintero, 1940). Presently, syphilis causes most infectious stillbirths in developing countries where it is prevalent, with 40 percent of fetuses dying in utero when maternal syphilis is left untreated (Goldenberg et al., 2010). PHUs combated pulmonary TB, which is associated with an elevated risk of perinatal death (Jana et al., 1994). Finally, PHUs possibly reduced morbidity through the anti-malarial measures described above. *Plasmodium falciparum* and *Plasmodium*

*vivax* malaria in pregnancy increase stillbirth risk, and both strains were prevalent (Moore et al., 2017).

### 2.3 Mortality Data and Health Unit Openings

I document the rollout of PHUs using annual reports of the Commissioner of Health and a history of PHUs written by a director of PHUs (Chaves, 1942). The dataset includes information on fiscal year of inauguration as well as a list of municipalities with PHUs and those municipalities with corresponding subunits, since the island was divided into districts consisting of one to four municipalities with shared health officers. In all but two municipalities, a PHU remained open through the end of the study period once it commenced operations. I assign the later date of inauguration to both municipalities but explore the robustness of results to dropping these municipalities in the appendix. Finally, there are discrepancies in the inauguration dates for a few municipalities, most importantly for several large cities and towns (San Juan, Ponce, Mayagüez, and Aguadilla). I rely on the earliest opening date found in official reports, but I demonstrate in the appendix that the regression results are robust to assigning inauguration dates based on Chaves (1942).

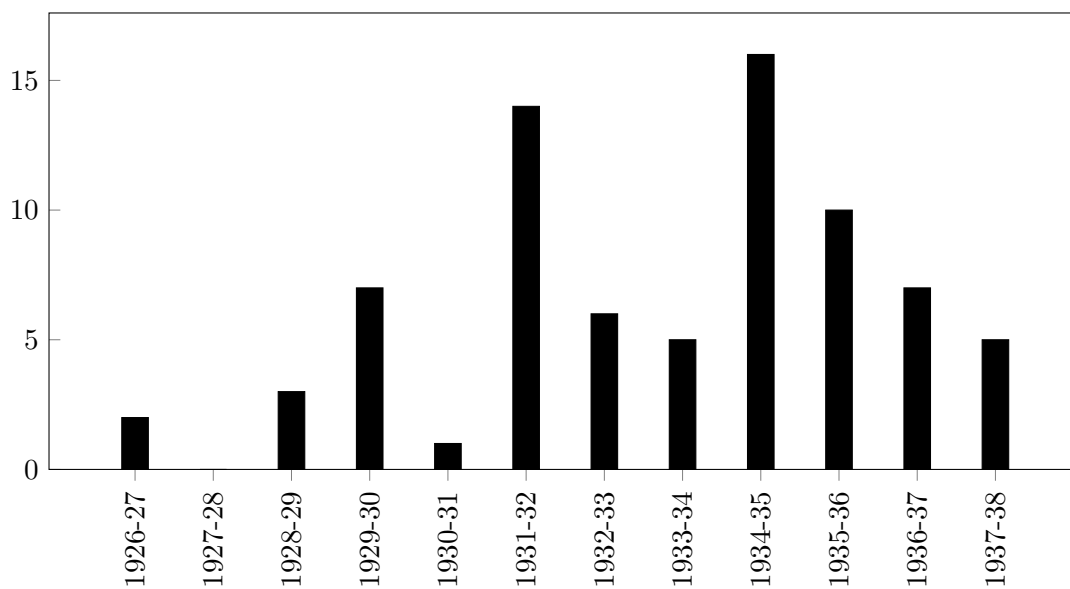
Figure 2.3 shows the number of PHUs openings by fiscal year and illustrates the rollout of PHUs over time and space from 1926 until all municipalities, save the sparsely populated island of Culebra (which was served by Vieques), had a unit in 1937. Río Piedras was selected as the site of a demonstration unit because it was considered characteristic of the northeastern coast of the island, contained urban and rural populations, allowed for easy communication with the central office in San Juan, was rapidly gaining in wealth and importance, was home to the university and various health facilities, and had municipal authorities willing to cooperate and fund the PHU (Martínez Rivera, 1943). The second unit opened in 1927 in Yabucoa, to represent the sugar region, followed in 1929 by Adjuntas, to represent the coffee region, and Cayey, to represent the tobacco region.<sup>2</sup> PHUs rapidly spread throughout the 1930s with constant pressure from municipal

---

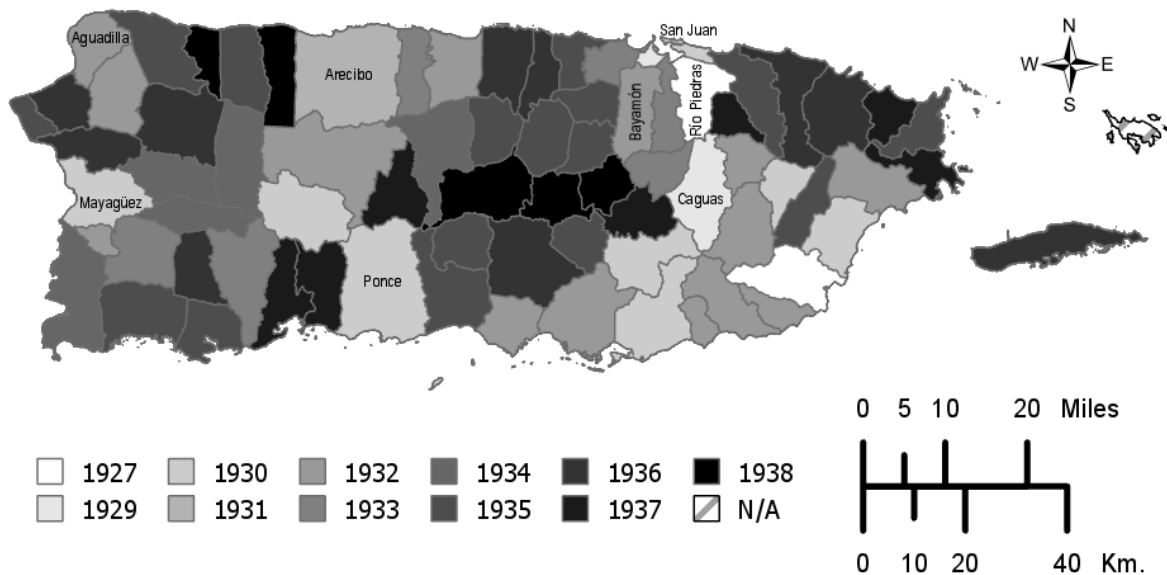
<sup>2</sup> PHUs also opened in fiscal year 1928-29 in Caguas, Cataño, and San Juan.

Figure 2.3: The Roll-out of Public Health Units

(a) Number of Public Health Units by Fiscal Year of Inauguration



(b) The Roll-out of Public Health Units by Municipality



Notes: These figures are based on the earliest opening date found in annual reports of the Commissioner of Health. Panel (a) does not include the opening of additional subunits within a municipality.

authorities (Chaves, 1942).

I link data on PHU operations to annual, municipal-level mortality data that I digitized from annual reports of the Commissioner of Health for 1923 to 1945.<sup>3</sup> Mortality is reported by fiscal year through 1929 and by calendar year thereafter. I construct mortality rates using population counts from the Census, linearly interpolating population for intercensal years. Annual data are available for all municipalities for the following years: overall mortality (1923-45), infant mortality (1924-45), stillbirths (1924-45), tuberculosis mortality (1923-45, except 1930), malaria mortality (1924-45, except 1927 and 1930), and maternal mortality (1924-1945, except 1927, 1929, and 1930).<sup>4</sup>

Mortality from tuberculosis in 1930 and from malaria in 1927 and 1930 are interpolated. All outcomes are measured by place of occurrence. The data include the entire population, as the municipalities (*municipios*) are analogous to counties in the United States and cover the whole territory.<sup>5</sup>

Figure 2.4 displays annual data on overall and infant mortality, which were regarded as reliable since all deaths were required by law to be registered (Hardy and Kramer, 1941).<sup>6</sup> A renewed decline in mortality rates coincided with the rollout of PHUs. Mortality from all causes returned to normal levels after the 1918-19 influenza pandemic, then fluctuated throughout the 1920s and early 1930s, with an uptick in 1928 resulting from Hurricane San Felipe. From 1933 forward, there was a general decline in overall mortality, from 22.2 in 1933 to 14.2 in 1945—a decrease of 36 percent. The infant mortality rate followed a similar pattern, averaging around 150 deaths before age one per 1,000 live births throughout the 1920s before starting a general decline in the early 1930s.

---

<sup>3</sup> Missing years were filled with monthly data from the *Porto Rico Review of Public Health and Tropical Medicine* or quarterly data from the *Puerto Rico Health Bulletin*.

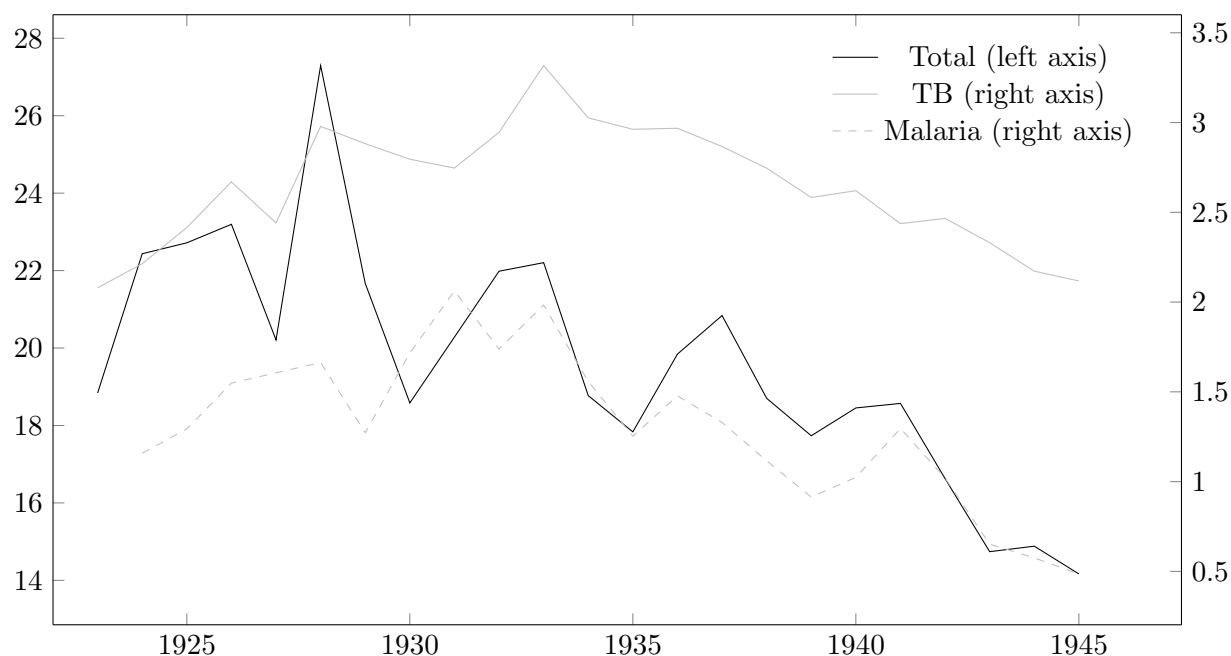
<sup>4</sup> Other forms of mortality are not frequently reported at the municipal-level in official reports, so I cannot use such outcomes in placebo tests. Instead, in the appendix I show that the estimated effect of PHUs is much more negative than would be expected from a placebo intervention in which treatment timing is randomly assigned.

<sup>5</sup> I exclude the island municipality of Culebra, due to missing data and its small population (e.g. 847 in 1930), and assign the same mortality rates for Bayamón and Cataño from 1923 to 1926 because Cataño split off of Bayamón in 1927 and hence did not report its own data in years prior.

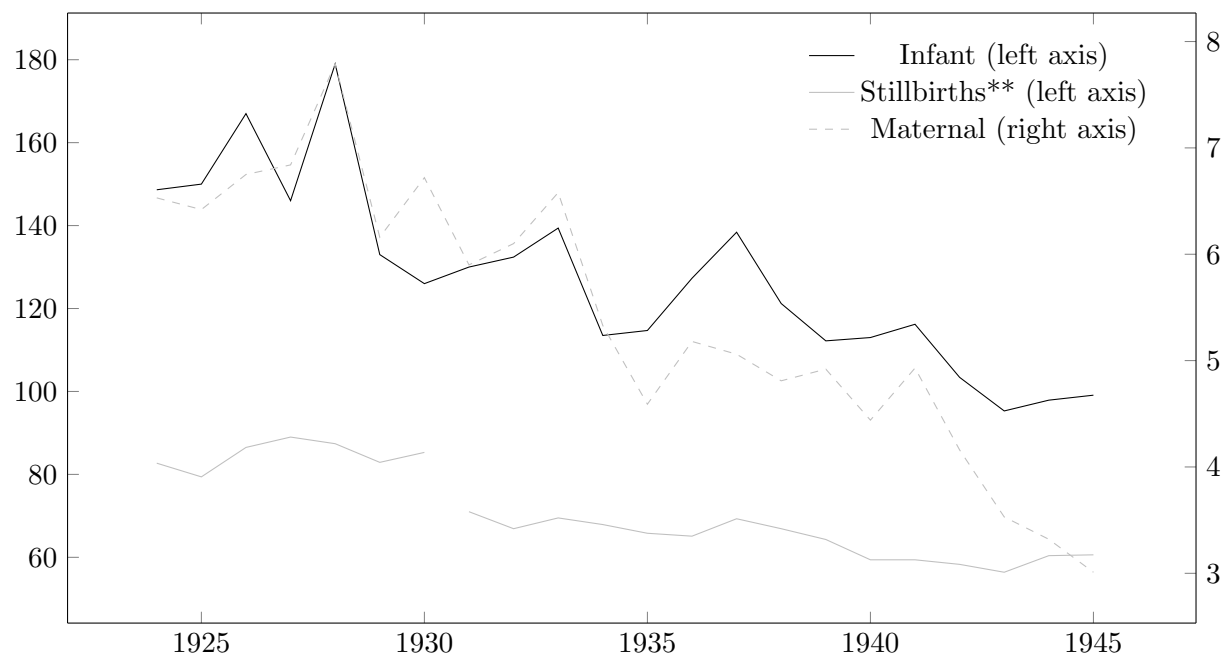
<sup>6</sup> The completeness of death registration stemmed from the tendency to bury the dead in cemeteries, resulting from the accessibility of cemeteries and the customs and religious beliefs of the vast majority of the population (Janer, 1945; Rodríguez Pastor and Janer, 1953; Vázquez Calzada, 1984). Consistent with its reputation for reliable death records, Puerto Rico was immediately admitted to the US Death Registration Area after its vital statistics system was centralized in 1931.

Figure 2.4: Mortality Rates, 1923-1945

(a) Deaths per 1,000 inhabitants



(b) Deaths per 1,000 births\*



Notes: Data are reported by fiscal year (e.g. 1924-25 corresponds to 1924) before 1930.

\*The infant mortality rate only includes live births.

\*\*Beginning in 1930, stillbirths were registered starting in the fifth month of pregnancy.

Sources: Annual reports of the Commissioner of Health of Puerto Rico, *Puerto Rico Review of Public Health and Tropical Medicine*, and the *Puerto Rico Health Bulletin*.

Between 1933 and 1943, the infant mortality rate fell by 32 percent, from 139.4 to 95.3. Similarly, the stillbirth rate hovered around 85 per 1,000 total births through 1930, dropped mechanically to 71.0 in 1931 after a new law required stillbirths to be registered only if the fetus had advanced to the fifth month of uterogestation, then declined to a low of 56.2 in 1943, or a decrease of 21 percent following the new registration law. Lastly, maternal mortality stayed around 6.8 deaths per 1,000 live births in the 1920s, fell to 5.9 in 1931 and below 5.0 by the late 1930s—around the time that sulfa drugs became widely available (Jayachandran et al., 2010; Thomasson and Treber, 2008)—and plummeted to 3.0 in 1945, a reduction of 57 percent.

Annual data on mortality rates for specific diseases are less reliable, especially before the creation of the Bureau of Vital Statistics in 1931 centralized and standardized data collection. Physicians often had to guess the cause of death in the absence of sound diagnostic practices and frequently signed death certificates without ever having seen the deceased (Commissioner of Health, 1926). Still, in Figure 2.4 I present mortality rates specific to TB and malaria, two of the deadliest diseases at the time and diseases which were targeted by PHUs. TB mortality exhibited an upward trend over the first half of the sample, rising by 62 percent between 1923 and 1932, but fell steadily thereafter and decreased by 34 percent by 1945.<sup>7</sup> Malaria mortality likewise increased, from 1.16 deaths per 1,000 inhabitants in 1923 to 2.06 deaths in 1931. Mortality from malaria then decreased by 76 percent through 1945, with much of the decline occurring during World War II, when the United States escalated its anti-malarial campaign to protect troops on the island (Miranda Franco and Casta Vélez, 1997). The apparent increases in deaths from TB and malaria up to 1931 are likely due to improvement in cause of death certification rather than the spread of those diseases.

Measurement error for mortality rates should be of little concern because it likely works against finding that PHUs decreased mortality. With classical measurement error, regression coefficients will be unbiased but standard errors will be inflated. Alternatively, it is conceivable that the underreporting of deaths declined after PHU openings. The revision and centralization

---

<sup>7</sup> A hurricane contributed to the spike in TB mortality in 1933, so I ignore this year in the discussion (Rodríguez Pastor and Janer, 1953).

of the vital statistics system in 1931 may have had a similar impact (Janer, 1958). In either case, the opening of a public health unit would artificially increase mortality rates and attenuate the estimated effect of PHUs.<sup>8</sup>

## 2.4 Event Study Framework

My empirical strategy exploits temporal variation in the establishment of PHUs to quantify their effect on mortality rates. I start with the following event-study specification:

$$\ln(Mortality_{mt}) = \theta_m + d_t + \sum_{k \neq -1} \gamma_k PHU_{mk} + \varepsilon_{mt} \quad (2.1)$$

where  $Mortality_{mt}$  is one of five mortality rates (overall mortality, infant mortality, TB mortality, malaria mortality, or stillbirths) in municipality  $m$  in year  $t$ .<sup>9</sup> Maternal mortality is not included in the event study because of consecutive years of missing data but is considered in a difference-in-differences specification in Section 2.5.3. Overall mortality rates are the number of deaths, excluding stillbirths, per 1,000 inhabitants; TB and malaria mortality rates are the number of deaths per 100,000 inhabitants; the infant mortality rate is the number of deaths, excluding stillbirths, before age 1 per 1,000 live births; and stillbirth rates are the number of stillbirths per 1,000 total births.  $\theta_m$  are municipality fixed effects to control for any time-invariant municipal characteristics correlated with the outcome variable, while  $d_t$  are year fixed effects to control for time-varying, territory-wide characteristics. Standard errors are clustered by municipality to allow for arbitrary correlation of errors within municipalities. Standard errors adjusted for spatial autocorrelation, regardless of the distance cutoff or time lag, are generally smaller than clustered standard errors, so they are relegated to Appendix A.1.10 (Colella et al., 2019; Conley, 1999). Observations are weighted by municipal population in 1925, just before the intervention.

<sup>8</sup> Births in Puerto Rico were underreported, so it is plausible that published infant mortality rates decreased over time as birth registration improved. However, in the appendix, I show that the results from the event study are nearly identical if the infant mortality rate is replaced with the ratio of infant deaths to population.

<sup>9</sup> TB and malaria mortality are coded as  $\ln(1 + Mortality_{mt})$ , where  $Mortality_{mt}$  is measured per 100,000 inhabitants, since some values of  $Mortality_{mt}$  are equal to zero. Results are robust to coding the dependent variable as  $asinh(Mortality_{mt})$ . Two observations report zero stillbirths, and I compute the stillbirth rate as if there were one stillbirth in both cases.

$PHU_{mk}$  is the event study indicator variable equal to one if the PHU in municipality  $m$  is  $k$  years away from its inaugural year ( $k = 0$ ) and zero in all other years.<sup>10</sup> Treatment is defined as being served by a PHU irrespective of the location of the PHU headquarters. In other words, treatment is assumed to be equivalent for municipalities with a unit and those with a subunit, or branch. Official reports offer no indication that municipalities with subunits were treated differently. The coefficients on  $PHU_{mk}$  characterize the effects of PHUs, relative to the year prior to PHU inauguration,  $k = -1$ , which is omitted.

I balance the panel by including leads and lags for all event-time periods for which there is an observation from each municipality. Observations more than four years (or three years for infant mortality, malaria mortality, and stillbirths) before or more than seven years after the establishment of a PHU are captured by “before” and “after” dummies. The binning of endpoints allows for identification of secular time trends and dynamic treatment effects when all units are eventually treated (Schmidheiny and Siegloch, 2020). The coefficients  $\gamma_k$  for  $k < -1$  reveal the changes in the mortality rate in the four (three) years preceding the opening of its PHU, after controlling for aggregate trends and time-invariant municipality characteristics that affect mortality, and allow me to directly assess whether the timing of PHUs is exogenous to trends or shocks in mortality preceding inauguration. The parameters of interest are  $\gamma_k$  for  $k \geq 0$ , which measure the effect of PHUs on mortality in each year for seven years following the opening of a PHU.

Identification of the average treatment effect requires that mortality in treated municipalities would have paralleled mortality in untreated municipalities if not for PHUs. If municipalities with the steepest (flattest) decline in mortality in the pre-period also were the first to open PHUs, the estimated effect of PHUs would be overstated (understated). This form of endogenous treatment, however, is unlikely because the earliest adopters were municipalities handpicked as test cases to

---

<sup>10</sup> Once again, mortality rates are reported by fiscal year through 1929 and by calendar year thereafter, but PHU treatment is reported by fiscal year over the entire sample. Mortality rates for fiscal year  $t/t + 1$  are coded as Year  $t$ , but a PHU opening in  $t/t + 1$  is coded as opening in  $t + 1$ . This coding scheme means that, for data from 1927 to 1929, municipalities in Event Year 0 were treated for the entire year and had already been treated between 0 and 12 months in the previous year, depending on the month of PHU opening within the fiscal year. For data from 1930 to 1945, municipalities in Event Year 0 were treated at least 6 months and up to 12 months that year and had already been treated between 0 and 6 months in the previous year.

represent the various regions of the island (see Section 2.2). After 1932, the Health Department sought to extend PHUs to all municipalities as rapidly as possible by branching out from existing units, so treatment timing over the latter half of the sample is quasi-random (Chaves, 1942; Martínez Rivera, 1943). The only other factor determining program adoption discussed in any government report is municipal budgets, such that the larger cities and towns were among the first to opt in. Even so, this only applies to the earliest units, as funding from the Puerto Rico Child Health Committee expedited the spread of PHUs in the 1930s and eventually municipalities “scarcely contributed” to the funding of PHUs (Chaves, 1937a, p. 26). Larger cities and towns tended to have higher pre-treatment mortality rates, but selection on levels does not bias results; instead, estimates of treatment effects will be biased only if there is selection on trends, which is unlikely given that public health work developed “slowly and gradually” until 1926 (Chaves, 1937a, p. 27) and there were no striking changes in general or infant mortality between 1910 and 1930 (Vázquez Calzada, 1978, 1984). Still, to capture potential diverging trends across municipalities, in the appendix I present specifications that include municipality-specific linear trends; an indicator variable equal to one if a municipality has a public water system and zero otherwise; and controls for the fractions of the population that were female, nonwhite, younger than 5, and literate, linearly interpolated between censuses.

A related threat to identification is that municipalities were treated in response to a local mortality shock, such that the estimated effect of PHUs may capture mean reversion. This form of endogenous treatment, too, is unlikely in that there is no indication in any government report that areas with worsening disease environments were prioritized. In any case, an advantage of the event study framework is that it allows for direct observation of mortality trends preceding the opening of a PHU, and there is no evidence of pre-trends or mortality shocks preceding treatment.

## 2.5 Public Health Units and Mortality

### 2.5.1 Baseline Results

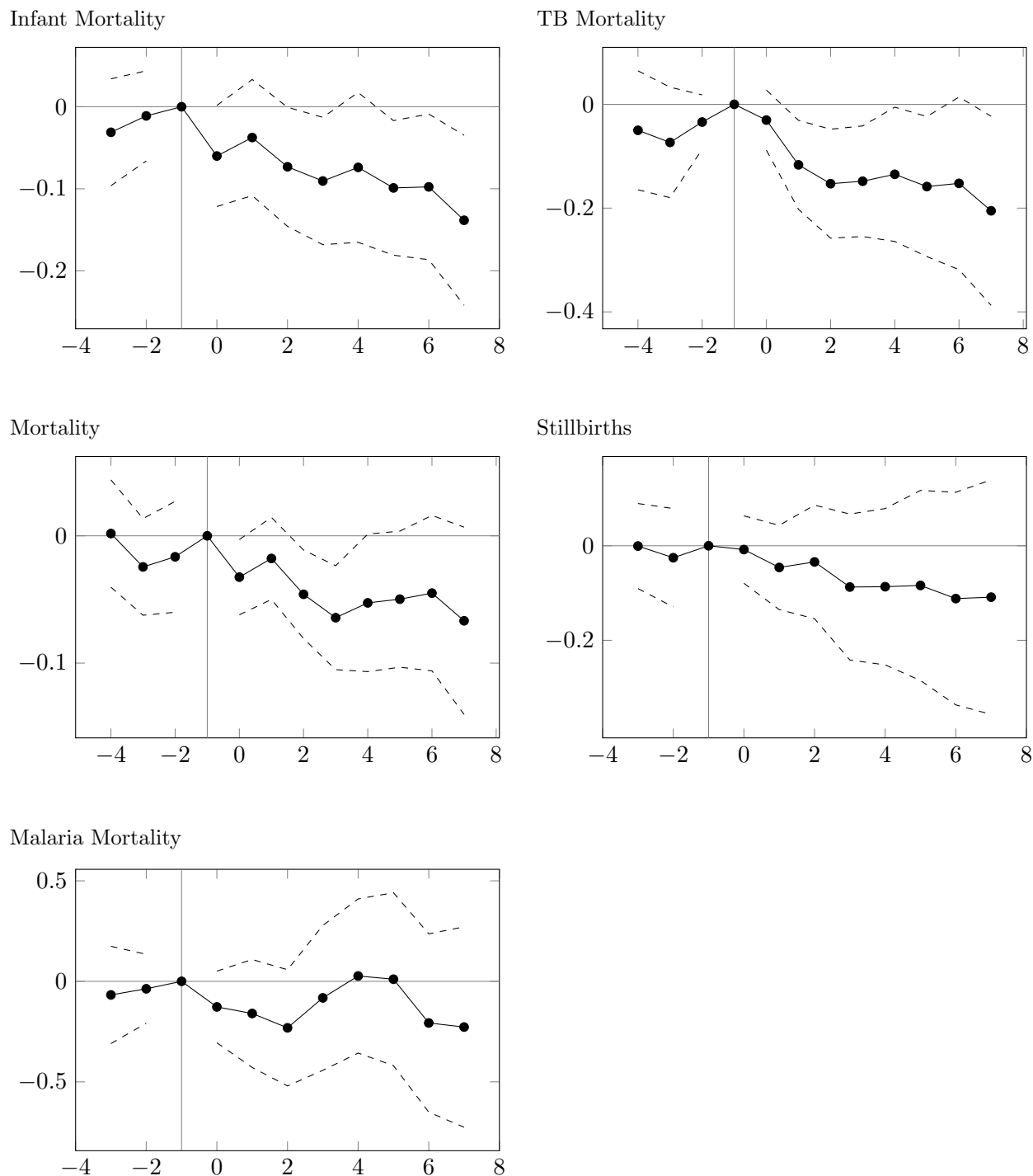
The results from the baseline event study specification are presented in Figure 2.5. There is neither a discernible trend nor any spike in any outcome preceding PHU entry that is not captured by controls. All estimates of  $\gamma_k$  prior to treatment are statistically indistinguishable from zero. This evidence strengthens the validity of the empirical strategy, confirming that the estimated treatment effects do not capture the mere continuation of falling mortality or mean reversion after a health shock. The absence of pre-trends lends credence to, but cannot prove, the assumption that mortality rates would have followed parallel trends in treated and untreated municipalities if not for PHUs.

The two health outcomes prioritized by PHUs, infant and TB mortality, significantly declined following program inauguration. Infant mortality declined by six percent (95% CI: [-11.4%, 0.2%]) in the first year, significant at the 10 percent level, and the effect of PHUs grew over the next several years. After two years, the estimated treatment effect is typically significant at the 5 percent level. Seven years after opening, PHUs reduced infant mortality by 13 percent ([-21.5%, -3.4%]).<sup>11</sup> The effect of PHUs on TB mortality was delayed but large: TB mortality dropped by 11 percent ([-18.4%, -3.3%]) after one year, and between 13 and 19 percent in subsequent years. The estimated treatment effect is statistically significant at the 1 percent level in event-years 1 to 3 and at the 5 or 10 percent level in event-years 4 to 7. The effect is comparable to that of TB dispensaries in Denmark, where TB mortality declined by around 19 percent and began to fall 17 to 19 months after the introduction of a dispensary (Hansen et al., 2020), consistent with an average time to death from notification of pulmonary TB in that era of 14.3 months for males and 12.9 months for females (Lindhardt, 1939). Unsurprisingly, general mortality likewise dropped following the opening of a PHU, by three percent ([-6.0%, -0.3%]) in the first year and by more than six percent ([-13.1%, 0.1%]) in event-year 7. The coefficients are usually significant at the 1 or 5 percent level

---

<sup>11</sup> The drop in infant mortality is not the result of changes in fertility, as the birthrate does not respond to the rollout of PHUs (see Appendix A.1.3).

Figure 2.5: The Effect of PHUs on Mortality: Baseline Specification



Notes: Dependent variables are log mortality rates. Dashed lines are 95 percent confidence intervals based on standard errors clustered by municipality. Observations are weighted by population. The event-year prior to inauguration is omitted. Event-years outside of the window are captured by “before” and “after” dummies. Sample includes 76 municipalities and 1,748 municipality-year observations for overall and TB mortality and 1,672 for infant and malaria mortality and stillbirths.

in early years and at the 10 percent level in later years.

There is also some evidence that PHUs reduced stillbirths. The estimated effect of PHUs is equal to zero in Event Year 0 but turns negative thereafter. PHUs are estimated to have brought down the stillbirth rate by four percent ( $[-12.6\%, 4.4\%]$ ) after one year and by 11 percent ( $[-28.6\%, 12.0\%]$ ) after six years, but never is the estimated effect close to statistical significance at conventional levels. Nonetheless, the consistently negative effect is noteworthy given the presumed high degree of measurement error in the dependent variable, as there is no universally accepted definition for when a fetal death is called a stillbirth versus a miscarriage (Tavares Da Silva et al., 2016). Moreover, it is not surprising that PHUs are found to have no negative effect on stillbirths until Year 1 even though there is a negative effect on general and infant mortality in Year 0. Whereas improved health services can have an immediate effect on the living, their effect on the unborn would materialize later.

Lastly, PHUs did not significantly affect malaria mortality. Although malaria mortality appears to fall after PHU inauguration, the decline is temporary, statistically insignificant, and not robust to alternative specifications, discussed in detail below. The historical record is clear that PHUs were not successful in combating malaria because they lacked sufficient funding and a trained field force (Arbona, 1942). The Annual Report of the Governor (1938, p. 34), in discussing anti-malarial efforts, states, “Because of limited funds it has not been possible to carry out permanent mosquito control measures on a sufficiently large scale to obtain measurable results.” Therefore, testing the effect of PHUs on malaria essentially is a placebo test; that I find no effect on malaria mortality strengthens the validity of the results for the outcomes which were prioritized by PHUs. This is important since no other forms of mortality are frequently included at the municipal-level in official reports.

### 2.5.2 Alternative Specifications

The specification above is standard in the literature, but binning the endpoints of the event window with “before” and “after” dummies may inadvertently bias estimates of  $\gamma_k$  (Borusyak and

Jaravel, 2018). For example, if long-run treatment effects are not constant, then the estimated effect for event-years after the event window will be unreliable. Through the wrong choice of unit and time fixed effects, the short-run effects will be biased.

I follow Borusyak and Jaravel (2018) to resolve this issue by imposing no restrictions on the dynamics of treatment effects. Terms corresponding to the year prior to PHU inauguration,  $k_1 = -1$ , and the earliest event-year in the sample,  $k_2$ , are dropped to be able to identify non-linear pre-trends.<sup>12</sup> The results from this specification appear in Figure 2.6 and are similar to the results of the baseline specification. Treatment effects are not estimated efficiently, so these results should be used only to evaluate pre-trends. Visual inspection of graphs for all outcomes offers no indication of trends preceding PHU openings, and coefficients before year 0 are generally quite close to zero. An  $F$ -test of the hypothesis that all leads of the treatment indicators are equal to zero renders the following  $p$ -values: 0.06, infant mortality; 0.23, TB mortality; 0.06, general mortality; 0.28, stillbirths; and 0.30, malaria mortality. Therefore, there is some concern about pre-trends, though in the case of infant mortality the low  $p$ -value is the product of a large spike 12 years prior to treatment, for which only 22 municipalities have data, and jumps up to  $p = 0.42$  when that event-year is excluded from the  $F$ -test. Stillbirths, likewise, spike downwards 13 and 12 years prior to PHUs before flattening out with coefficients near zero.

The two-way fixed effects models presented above are not robust to heterogeneity in treatment effects across calendar years or municipalities. Some cohorts may be assigned negative weights when estimating dynamic treatment effects, biasing the event study coefficients. Additionally, tests of pre-trends are invalid when treatment effects are heterogeneous (Abraham and Sun, forthcoming). Assuming homogeneous treatment effects may be unrealistic, as PHUs might have had different effects in municipalities depending on preexisting local health services, for example, or in certain calendar years due to macroeconomic conditions.

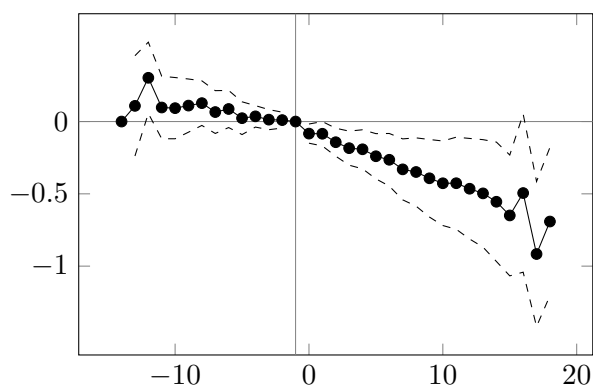
I present results from an estimator proposed by de Chaisemartin and D’Haultfoeuille (2020)

---

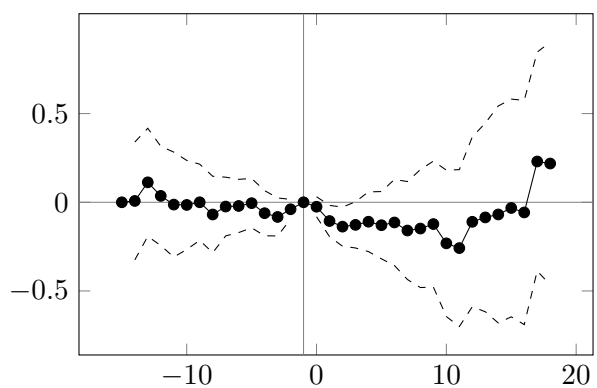
<sup>12</sup> The formal test of pre-trends is invariant to the choice of  $k_1$  and  $k_2$ . Dropping periods far away from each other decreases the likelihood of observing a linear pre-trend (which is impossible to detect) and distracting attention from nonlinearities in pre-trends.

Figure 2.6: The Effect of PHUs on Mortality: Dynamic Specification

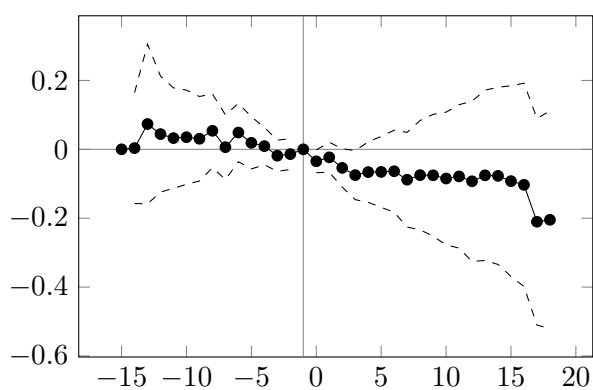
Infant Mortality



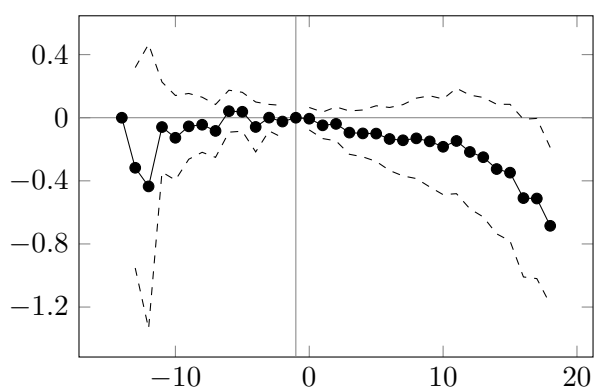
TB Mortality



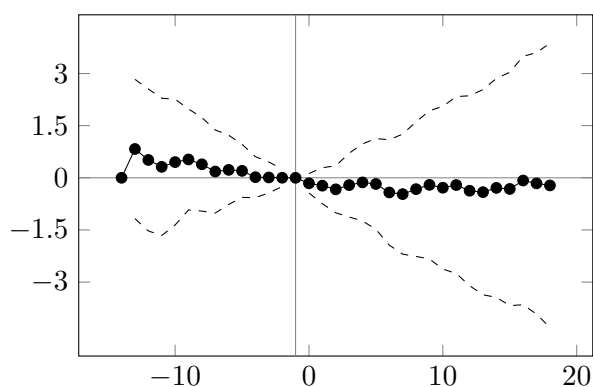
Mortality



Stillbirths



Malaria Mortality



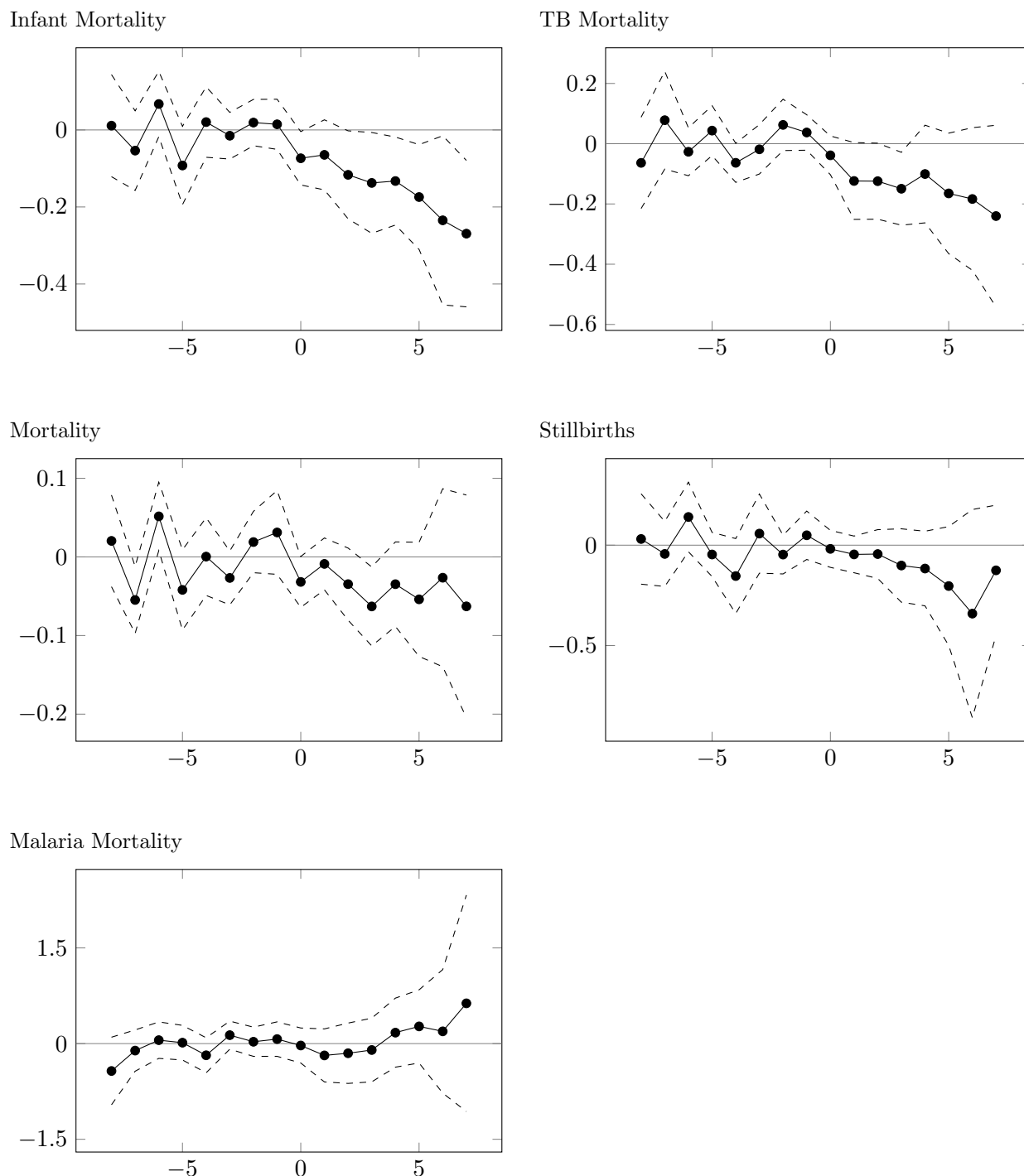
Notes: Dependent variables are log mortality rates. Dashed lines are 95 percent confidence intervals based on standard errors clustered by municipality. Observations are weighted by population. The event-year prior to inauguration and the earliest event-year are omitted. Sample includes 76 municipalities and 1,748 municipality-year observations for overall and TB mortality and 1,672 for infant and malaria mortality and stillbirths.

that is valid even if the treatment effect is heterogeneous over time or across municipalities. The estimand identifies the effect of the treatment in the groups that switch treatment, at the time when they switch, by comparing them to those that do not become treated at that time. For each event-time period, the estimand is the average of the instantaneous treatment effect at the time when a municipality starts receiving the treatment, across all municipalities. The estimator relies on a common trends assumption.

The results from this estimator appear in Figure 2.7 and are similar to those presented above. There is no evidence of pre-trends for infant mortality, TB mortality, malaria mortality, or stillbirths. The placebo coefficients for general mortality are large for event-years at least five years prior to treatment, but even so there is no clear trend. The two outcomes prioritized by PHUs, infant and TB mortality, decline substantially after treatment. Infant mortality falls by seven percent (95% CI: [-13.3%, -0.4%]) in the first year and by 24 percent ([-36.8%, -7.6%]) after seven years; all post-treatment coefficients other than event-year 1 are significant at the 1 or 5 percent levels. TB mortality declines by 12 percent ([-22.2%, 0.4%]) after one year of treatment and by about 20 percent after several years; post-treatment coefficients are significant at the 5 or 10 percent level in most years. Stillbirths also drop, by four percent ([-12.8%, 4.6%]) after one year and by more than 10 percent in subsequent years, although the treatment effects are never statistically significant. Malaria mortality exhibits no trend, falling briefly after program inauguration before increasing for several years; the event study coefficients for malaria are never significant. Finally, general mortality does clearly decline after program inauguration—a trend overshadowed by unstable pre-treatment coefficients—but the effect is only significant at the 10 percent level in event-year 0 and at the 5 percent level in event-year 3. The treatment effect varies from three to six percent in the first several years with a PHU.

To summarize, all specifications shown in this paper produce similar results and support the conclusion that PHUs reduced the incidence of infant mortality and TB mortality and probably contributed to the decline of stillbirths. The point estimates from this section indicate that PHUs were responsible for more than half of the overall reduction in infant and TB mortality through

Figure 2.7: The Effect of PHUs on Mortality: de Chaisemartin and D'Haultfoeuille Estimator



Notes: Dependent variables are log mortality rates. Coefficients are derived from the difference-in-differences estimator from de Chaisemartin and D'Haultfoeuille (2020). Dashed lines are 95 percent confidence intervals based on standard errors clustered by municipality. Observations are weighted by population. Sample includes 76 municipalities and 1,748 municipality-year observations for overall and TB mortality and 1,672 for infant and malaria mortality and stillbirths.

1945, lowering each by more than 20 percent after several years. Counterfactual estimates derived from the baseline specification suggest that PHUs prevented 1,136 infant deaths and 1,105 deaths from TB in 1940 alone.<sup>13</sup>

Oddly, the results for general mortality are less straightforward despite strong evidence of reductions in infant and TB mortality. Estimated treatment effects for general mortality are noisier, and estimated pre-treatment coefficients are not always close to zero despite displaying no trend. Still, each estimator points to a reduction of about six percent in general mortality, or nearly one-fifth of the total decline in mortality between 1923 and 1945 or one-third of the decline before US entry into World War II, during which the federal government launched an anti-malarial campaign to protect troops on the island and saved at least a thousand lives annually by the war's end (Miranda Franco and Casta Vélez, 1997).

The estimated treatment effects on stillbirths are statistically insignificant but consistently large and negative, so the evidence is suggestive that PHUs reduced stillbirths given substantial measurement error in the dependent variable. Imprecise estimates and the redefinition of stillbirths in 1931 make it difficult to determine precisely the share of the decline attributable to PHUs, but PHUs seem to be the most important factor since the stillbirth rate fell by less than 20 percent after 1931.

Finally, the evidence is clear that PHUs did not reduce malaria mortality. This result strengthens the results for the other outcomes because we should expect a null result. The historical record is unanimous in concluding that PHUs did not meaningfully contribute to the eradication of malaria, so malaria mortality in my analysis essentially serves as a placebo outcome in lieu of other cause-of-death data that were not frequently reported by the Department of Health. In sum, the event study demonstrates that PHUs succeeded in reducing the outcomes that they targeted—principally, infant and TB mortality—and had no effect on an outcome that they lacked

---

<sup>13</sup> Counterfactual estimates are derived by comparing estimated mortality with and without (i.e. setting all treatment dummies to zero) treatment. Using 95-percent confidence intervals for the estimated effects of PHUs to derive lower and upper bounds, these estimates range from 105 to 1,977 infant deaths and 0 to 2,480 TB deaths prevented.

the human and financial resources to deal with—namely, malaria mortality.

### **2.5.3 Nurses, Sanitary Inspectors, and Assistant Midwives**

PHUs carried out a wide range of activities, making it difficult to know which efforts were responsible for reducing mortality. Official reports do not contain yearly, municipal-level information on the personnel employed or services provided by health units. In this section I use data on the full-time personnel employed by PHUs in 55 municipalities in 1935 to describe how the effect of PHUs varied with the number of nurses and sanitary inspectors per capita. I also incorporate data on the number of assistant midwives per capita in 1935-36 to comment on occupational licensing.

Exercises based on these data are descriptive and should be interpreted with caution for several reasons. First, it is plausible that these snapshots are not representative of municipal-level inputs over time. Second, personnel per capita may not be randomly assigned, and municipalities with more nurses, sanitary inspectors, or midwives may have dedicated more resources to complementary aspects of their public health programs that are not captured by these variables. Third, in the case of nurses and sanitary inspectors per capita, the analysis is based on a sample of already-treated municipalities.

To assess whether the effect of PHUs varied with the number of nurses or sanitary inspectors per capita, I run a regression with a full set of lags interacted with the number of nurses and inspectors per capita in 1935, respectively, as reported by Mountin et al. (1937). Fifty-five municipalities are included in the sample, and those municipalities without a PHU in 1935 are excluded. The number of public health nurses per 10,000 inhabitants ranged from 0.3 to 2.6, and the average municipality had 1.2 (median=1.0, s.d.=0.6). The number of sanitary inspectors per 10,000 inhabitants ranged from 0 to 4.8, and the average municipality had 0.9 (median=0.7, s.d.=0.7).

Table 2.1 reports coefficients that are linear combinations of coefficients from event-years 0 through 7. Municipalities with more nurses per capita experienced larger decreases in infant

Table 2.1: Changes in Mortality by Personnel Per Capita

<i>Dependent variables:</i>	IMR	TB	Mortality	Stillbirths
PHU	-0.06 (0.06)	-0.15 (0.15)	-0.05 (0.04)	-0.10 (0.11)
PHU×Nurses <sub>1935</sub>	-0.10** (0.05)	0.05 (0.07)	-0.03 (0.02)	-0.05 (0.05)
Observations	1,210	1,265	1,265	1,210

<i>Dependent variables:</i>	IMR	TB	Mortality	Stillbirths
PHU	-0.16** (0.07)	-0.12 (0.17)	-0.07 (0.04)	-0.13 (0.11)
PHU×Inspectors <sub>1935</sub>	0.00 (0.04)	0.02 (0.08)	-0.02 (0.02)	-0.02 (0.04)
Observations	1,210	1,265	1,265	1,210

<i>Dependent variables:</i>	IMR	TB	Mortality	Stillbirths	Maternal
PHU	-0.17*** (0.05)	-0.13* (0.07)	-0.06** (0.03)	-0.18* (0.10)	-0.21* (0.11)
1[t ≥ 1931] ×Midwives <sub>1935–36</sub>	-0.12*** (0.04)	-0.05 (0.06)	-0.08*** (0.02)	0.03 (0.08)	-0.43*** (0.14)
Observations	1,672	1,748	1,748	1,672	1,444

Notes: The dependent variables are log mortality rates. The coefficients are weighted least-squares estimates. Standard errors are clustered by municipality.  $Nurses_{m,1935}$  and  $Inspectors_{1935}$  represent the number of nurses and inspectors, respectively, per 10,000 inhabitants in 1935 in municipality  $m$ .  $Midwives_{m,1935-36}$  represents the number of licensed assistant midwives per 1,000 inhabitants in 1935-36. Coefficients on  $Nurses_{m,1935}$ ,  $Inspectors_{1935}$ ,  $Midwives_{m,1935-36}$  are estimated as a full set of lags interacted with the personnel variable. The coefficients reported are a linear combination of coefficients estimated separately for event-years 0 to 7. The number of midwives per capita is interacted with a dummy equal to one in all years starting with 1931, since the licensing of assistant midwives started that year. See Sections 2.5.2 and 2.5.3 for more details. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

mortality, such that one additional nurse per 10,000 inhabitants is associated with 10 percent lower infant mortality. There is no evidence that changes in any type of mortality varied by per

capita number of sanitary inspectors. These results suggest that the services of nurses may have been more important than those of sanitary inspectors for reducing infant mortality. Municipalities maintained a sanitation service prior to PHUs, so the null results for sanitary inspectors may reflect that PHUs merely continued these services (Mountin et al., 1937). These results are consistent with recent research on other historical contexts. Similar efforts led by nurses reduced infant mortality in the early 20th century in the United States (Moehling and Thomasson, 2014), Denmark (Wüst, 2012), Sweden (Bhalotra et al., 2017), and Norway (Bütikofer et al., 2019).

I run similar regressions with assistant midwives per capita as reported in the 1935-36 Commissioner of Health report. A 1931 law required midwives to be licensed by PHUs, but 4,653 documented births were attended by non-authorized persons in fiscal year 1935-36. The law applied to all municipalities, so regressions include an interaction of the number of midwives per capita and an indicator variable equal to one for all years starting in 1931. The number of assistant midwives per 1,000 inhabitants ranged from 0.9 to 3.6, and the average municipality had 1.9 (median=1.7, s.d.=0.6).

An additional midwife per 1,000 inhabitants is associated with a 12 percent drop in infant mortality. The number of midwives is also significantly negatively related to general mortality, though the magnitude of the coefficient is 40 percent smaller. Decreases in stillbirths and TB mortality, which are unrelated to the work of midwives, did not vary with the number of midwives.

These findings are consistent with recent research: licensed midwives reduced infant mortality modestly in the United States from 1900 to 1940 (Anderson et al., 2020a) and significantly in Sweden from 1881 to 1930 (Lazuka, 2018). In Puerto Rico, most of the decline in infant mortality resulted from a higher probability of survival in the first month of life (Wegman et al., 1942), and assistant midwives were trained in proper hygiene for childbirth and visited for days after birth to wash and monitor the mother and baby (Robert de Romeu, 1939).

Finally, I examine PHUs in relation to maternal mortality in a difference-in-differences specification since data are missing for 1923, 1927, 1929, and 1930.<sup>14</sup> An additional midwife

---

<sup>14</sup> I consider maternal mortality only in relation to midwives because the universal application of the licensure law

per 1,000 inhabitants is associated with 35 percent lower maternal mortality, significant at the 1 percent level ( $p = 0.004$ ). PHUs without the licensure law are associated with a 19 percent decline in maternal mortality ( $p = 0.07$ ), perhaps through the monitoring and education of mothers at prenatal clinics. However, a static model with a staggered rollout should be viewed with caution due to problems caused by negative weighting (de Chaisemartin and D’Haultfoeuille, 2020).

These results are consistent with trends in causes of maternal mortality. Deaths from complications of childbirth that could be affected by the work of midwives, such as puerperal sepsis, gradually declined after 1931, but deaths from puerperal hemorrhage, for which an assistant midwife could do little to help, stayed about the same (Robert de Romeu, 1939). These results are also supported by recent studies on other historical contexts. The licensing of midwives significantly reduced maternal mortality in the United States from 1900 to 1940 (Anderson et al., 2020a), Norway from 1887 to 1921 (Kotsadam et al., 2017), and Sweden from 1830 to 1894 (Pettersson-Lidbom, 2015).

## 2.6 Cost-Benefit Analysis

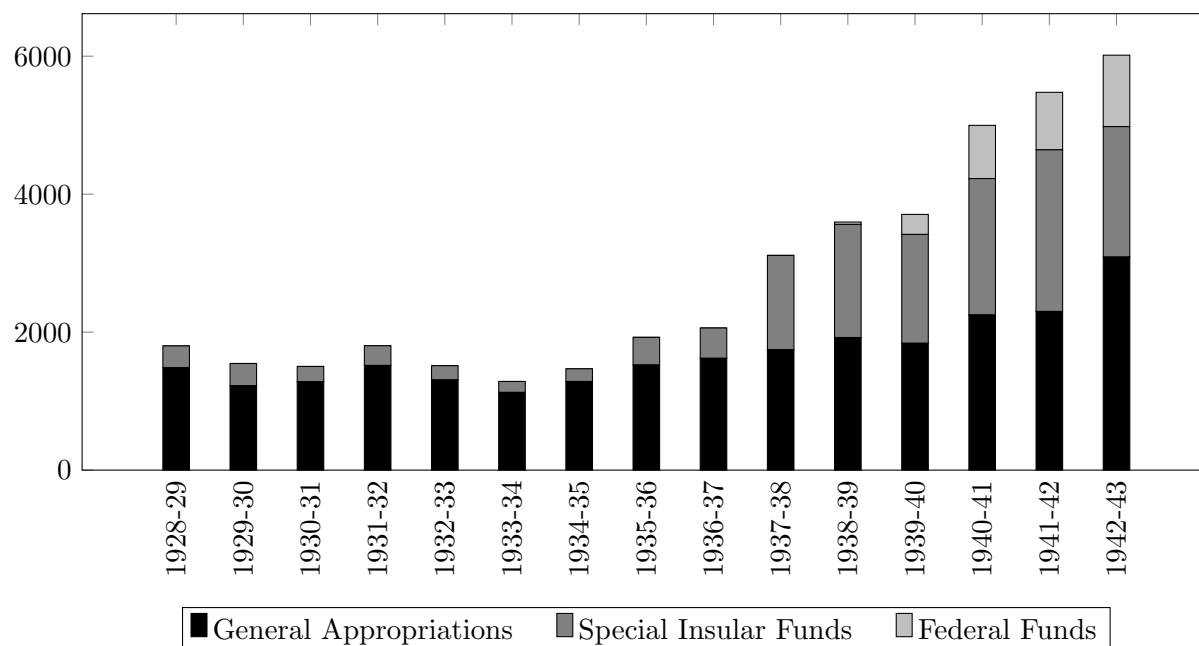
Remarkably, the reduction in mortality brought about by PHUs came at little additional cost to taxpayers. The financial resources commanded by the Department of Health did not markedly grow until after the rollout of PHUs (see Fig. 2.8). The constraints imposed by the Great Depression—including a 29 percent (\$518,109) budget cut from 1931-32 to 1933-34—prompted the department to simplify its organization and eliminate unnecessary expenses, such as by cutting most salaries by 10 percent (Commissioner of Health, 1934; Ramírez de Arellano, 1981). Yet, the department carried out, and even expanded, its ordinary activities. For instance, from 1931-32 to 1933-34, total and per capita costs fell at the insular psychiatric hospital, TB hospital, leper hospital, quarantine hospital, charity schools, blind asylum, institute for blind children, and the Ponce District Hospital, yet no institution severely cut back on its number of patients

---

allows me to flexibly observe the correlation between midwives and maternal mortality over time (see appendix). By contrast, a triple-differences result for, e.g., PHUs and nurses per capita, cannot be estimated flexibly in an event study model due to missing years of data.

(Commissioner of Health, 1939).

Figure 2.8: Resources of the Department of Health, 1928-29 to 1942-43  
(Thousands of dollars)



Source: *Report of the Commissioner of Health of Puerto Rico for the Fiscal Year 1942-43*

Appropriations for health units doubled throughout the 1930s and constituted about one-fifth of total insular health expenditures over the decade (see Table 2.2). Before World War II, expenditures on PHUs made by the Department of Health averaged \$0.25 per capita (\$4.61 in 2020 dollars). However, these numbers fail to account for contributions from municipalities, the Rockefeller Foundation, and external sources. The Rockefeller Foundation assigned a supervisor to the health units and contributed financially—for example, \$14,000 to PHUs in specific and \$72,000 to the Health Department in general in 1931-32—to the training of nurses and support of the PHUs (Commissioner of Health, 1932). Other notable contributions include \$25,000 from the US Public Health Service in 1934-35 toward the payment of salaries; about \$35,000 annually throughout the early 1930s from the Puerto Rico Child Health Committee, also to help cover salaries; and \$325,000 from the Puerto Rico Reconstruction Administration in 1937 and 1938 for the construction of 24

new buildings (Commissioner of Health, 1935, 1939).<sup>15</sup>

Table 2.2: Expenditures on PHUs by the Insular Department of Health, 1931-32 to 1938-39

Fiscal Year:	31-32	32-33	33-34	34-35	35-36	36-37	37-38	38-39
Appropriations (\$1000s)	253	236	228	306	377	385	454	519
Share of insular health costs	0.14	0.16	0.18	0.21	0.20	0.19	0.15	0.15
Municipalities covered	28	34	38	54	64	71	77	77
Population covered (1000s)	872	986	1,073	1,335	1,528	1,641	1,745	1,767
Per capita cost (\$)	0.29	0.24	0.21	0.23	0.25	0.23	0.26	0.29

Source: Puerto Rico Department of Health

I convert estimates of infant and TB deaths averted from Section 2.5 to dollars using estimates of the Value of Statistical Life (VSL) and the elasticity of the value of life with respect to per capita GNP for the United States in 1940 from Costa and Kahn (2004).<sup>16</sup> Life-years saved are estimated using life expectancy in the 1930s at various ages from Vázquez Calzada (1978) and the distribution of TB mortality by age from Rodríguez Pastor and Janer (1953). I derive the value of a life-year in Puerto Rico in 1940,  $VOLY_{1940,PR} = VOLY_{1940,US}(Y_{1940,PR}/Y_{1940,US})^\epsilon$ , where  $Y_t$  is GDP per capita in year  $t$  and  $\epsilon$  is 1.7, the upper bound estimate of the value of life with respect to per capita GNP. By necessity, I assume that preferences over life and income were the same in the United States and Puerto Rico and that income in the United States was five times the level in Puerto Rico (Devereux, 2019).<sup>17</sup> The resulting value of a life-year is \$2,151 in 2020 USD. I estimate total program costs by adding the expenditures of the Department of Health on PHUs in Table 2.2 with available information about contributions from external sources and an estimate of municipal contributions from Vilar Isern (1935). My estimate of costs overstates the marginal costs of PHUs, since PHUs subsumed preexisting sanitation services. Likewise, my estimate of benefits is conservative because life-years saved is based on life expectancy in the mid-1930s, during a time of rapid increases in life expectancy for all age groups; I do not include stillbirths averted due to

<sup>15</sup> The Department of Health also received federal funds for activities other than PHUs, such as \$431,500 for the construction of hospitals (Commissioner of Health, 1938).

<sup>16</sup> I impose a lower bound estimate of zero infant or TB deaths averted in any given year.

<sup>17</sup> VSL values are converted to 1940 dollars using the Bureau of Labor Statistics historical Consumer Price index (2020).

imprecise estimates or maternal deaths averted due to missing years of data; and the estimated number of infant lives saved is much smaller than would be derived from other specifications.

A back-of-the-envelope calculation indicates that the benefits of PHUs exceeded costs by a ratio of 13 to 1 from 1931-32 to 1938-39 (see Table 2.3). Using 95-percent confidence intervals for the estimated effects of PHUs to derive lower and upper bounds, the benefit-cost ratio ranges from 1:1 to 26:1. Puerto Rico spent about \$6,653 (2020 dollars) per life saved or \$168 per life-year saved. These rough estimates strongly suggest that the benefits of PHUs exceeded their costs.

Table 2.3: Cost-benefit Analysis, 1931/32-1938/39

	Estimate	Lower bound	Upper bound
Infant deaths averted	4,554	241	9,394
TB deaths averted	5,347	623	10,867
Life-years saved	393,107	31,591	805,821
Cost per averted death (2020 dollars)	6,653	76,269	3,251
Cost per life-year saved (2020 dollars)	168	2,085	82
Value of life year (2020 dollars)	2,151	2,151	2,151
Benefits (thousands of 2020 dollars)	845,376	67,936	1,732,918
Costs (thousands of 2020 dollars)	65,870	65,870	65,870
Rate of return (benefit-cost ratio)	13:1	1:1	26:1

Notes: Life-years saved are estimated using life expectancy in the 1930s at various ages from Vázquez Calzada (1978) and the distribution of TB mortality by age from Rodríguez Pastor and Janer (1953). Value of statistical life for Puerto Rico in 1940 is derived from estimates for the US in 1940 for 18- to 30-year-old men from Costa and Kahn (2004). Lower and upper bound estimates of deaths averted are based on the 95 percent confidence intervals of treatment effects by event-year from Section 2.5 and imposing a lower bound of zero on the number of lives saved in a year. All monetary values were converted to 2020 dollars using the Bureau of Labor Statistics Historical Consumer Price Index (2020).

## 2.7 Conclusion

This paper studies the effect of public health units on mortality during the early years of the mortality transition in Puerto Rico, the first outside of Europe and Western offshoots and one of most successful in world history. Using a novel dataset of annual, municipal-level vital statistics from 1923 to 1945, I exploit the staggered rollout in an event study framework, assuming

that municipalities would have followed parallel trends if not for PHUs. Health units accounted for most of the decline in infant and TB mortality and about one-fifth of the decline in general mortality through 1945. PHUs also reduced maternal mortality and stillbirths but had no effect on malaria mortality. While it is difficult to know which PHU services made a difference, the finding that reductions in infant and maternal mortality were steeper with more nurses and midwives per capita, but unrelated to the number of sanitary inspectors, suggests that efforts led by nurses (e.g. home visits and prenatal clinics) and the licensing of midwives may have been more important than the enforcement of sanitary regulations or improvements in rural sanitation.

This paper contributes to a debate about the relative importance of public health and economic growth in the global mortality transition. While not ruling out the importance of rising income, I demonstrate that public health units were instrumental in reducing mortality during the initial phase of the mortality transition in Puerto Rico. Of course, regionally high public health expenditures were made possible by consistent economic growth after 1898, and economic factors may have become more important after 1945 when the upward trends in both per capita income and life expectancy accelerated. Still, public health set in motion sustained decreases in mortality.

Bringing evidence from Latin America into the discussion is an important step in looking beyond Europe and Western offshoots. The study of Puerto Rico is possible because of its affiliation with the United States, which produced extensive administrative records and vital statistics at US standards. US interventions in Latin America in the 20th century may have generated comparable records, allowing for wider study of the region. Notably, Cuba and Panama received major public health investments during periods of American control.

Discussions of Latin American development tend to exclude Puerto Rico because its special relationship with the United States supposedly makes it unlike independent countries. Yet, after Puerto Rico, practically all countries of Latin America opened public health units, so this investigation provides a window into public health in Latin America more broadly (Bravo, 1958). Local health services in Puerto Rico and throughout much of Latin America were realized with the technical and financial aid of the Rockefeller Foundation and the US federal government. In any

case, most public health efforts in Puerto Rico were carried out by Puerto Ricans trained on the island, and health units were implemented not through new transfers from the mainland but partly by redistributing public health expenditures for greater effectiveness (Payne, 1942). The success of the program, the emphasis on administrative reform, and the reliance on low technology but labor-intensive measures makes Puerto Rico a useful example for countries with limited resources.

## Chapter 3

### Colonial Roads and Regional Inequality

#### 3.1 Introduction

Transport infrastructure is widely considered fundamental to economic development (Berg et al., 2017). Roads facilitate trade and provide access to education and health care. The location of infrastructure also influences the spatial distribution of economic activity: roads and railways promote local economic development in the short run, and the benefits may persist in the long run even after the initial advantages of the investments, or the infrastructure itself, have disappeared (Berger and Enflo, 2017; Bertazzini, forthcoming; Jedwab and Moradi, 2016; Jedwab et al., 2015; Okoye et al., 2019; Roessler et al., 2020). Counterintuitively, investments in transport infrastructure can exacerbate regional inequality (Baum-Snow et al., 2020), although the evidence is mixed (Fageda and Olivieri, 2019).

In this paper, I consider road-building, local economic development, and regional inequality by examining the ambitious effort in the early 20th century to connect all cities and towns in Puerto Rico with roads. After the United States annexed Puerto Rico in 1898, American officials considered roads “the crying need of the island” (Governor of Porto Rico, 1901, p. 73) and “the subject of the greatest importance” in its development (Carroll, 1899, p. 40). The United States prioritized road-building in the mountainous interior, where transportation facilities were practically nonexistent (see Davis, 1900, p. 373). Soon roads connected all areas of the island, forming one of the densest networks in the world (Clark, 1930; World Bank, 2011); at present, 99.4 percent of the island area is located within one kilometer of the nearest public road (Grau et al.,

2004).<sup>1</sup>

Using newly digitized historical maps of the transport network and district-level population data, as well as several measures of modern development, I consider how historical investments in roads shaped the spatial distribution of economic activity in the short and long run. I conduct analysis at the level of the municipality as well as the more granular 2.5 arc-minute resolution (approximately 4.5×4.5 km. squares) to exploit the full range of modern data and evaluate the robustness of the main results to different ways of measuring access to the transport network. First, I show that the negative relationship between distance to roads in 1899 and population density in year  $t$  increases in magnitude throughout the early 20th century. In 1899, roads connected the largest cities and towns, but proximity to roads does not strongly predict population density away from the network nodes. After 1899, the population increasingly clustered near old roads built under Spanish rule—and modernized under early US rule—despite the rapid construction of roads elsewhere. Areas near old Spanish roads became, and remain, richer and more densely populated than other areas, including those near American roads. Hence, the descriptive evidence reveals that investments made under early US rule not only failed to develop areas previously lacking access to the transport network but, consistent with theory (Krugman, 1991) and empirical evidence in other settings (Baum-Snow et al., 2020), plausibly led to greater spatial inequality. As observed by Jackson (2020, p. 124), US highways “seemed to facilitate the flow of labor and capital from the coffee uplands to the sugar-growing lowlands, and not the reverse, as originally intended. . .”

Second, I employ an instrumental variables strategy at both levels of analysis to identify the causal effect of historical roads on local development. This approach exploits plausibly exogenous variation in the location of roads and addresses concerns about reverse causality (e.g., Berger and Enflo, 2017; Jedwab and Storeygard, 2018). At the grid-level, I instrument distance to roads in 1899 with distance to straight lines connecting the defunct seats of Spanish administration, as reflected by the nodes of the telegraph network authorized in 1866. I exclude areas near the nodes

---

<sup>1</sup> Roads in Puerto Rico are not merely extensive but also of high quality: the World Economic Forum (2015) ranks Puerto Rico 29th of 141 countries in terms of quality of roads, just ahead of the United Kingdom.

because they were endogenously chosen and assess the impact of roads in areas which happened to gain access to the network by virtue of lying between major cities. At the municipal-level, I instrument market access in 1899 with the average travel time to all cities in other departments. By focusing on travel times to faraway municipalities, I aim to capture variation attributable to roads not nearby and thus less likely to be driven by local productivity. Travel times also do not incorporate market size, which is more clearly endogenous to local productivity than the location of public roads.

I measure the long-run effect of roads on a wide range of modern outcomes: population density, nighttime light density, urban extent, per capita income, number of firms, number of factories, and an index of socioeconomic development. My results indicate that early investments in roads promoted local economic development, and this finding is robust to various specifications and controlling for distance to the coast, elevation, land gradient, rainfall, and crop suitability (for sugarcane, tobacco, and coffee), as well as several historical development indicators including population density, literacy, and agricultural employment.

These results suggest that the effect of roads varied with time: roads fostered local economic development, but their effect increased during the rise of commercial agriculture and external trade in the early 20th century. I use historical census data to show that municipalities with greater market access experienced faster population growth only after 1898. This finding parallels the experience of French Algeria, where railways had a positive effect on indigenous population growth only after cereal cultivation improved and trade policy changed such that there was sufficient economic life to render the network profitable (Buckwalter, 2019).

Subsequent US investments largely failed to redirect economic activity to remote areas. The entire island was integrated into the network within two or three decades of US annexation, yet modern outcomes are driven more by access to roads built prior to the 20th century, suggesting that early access to the transport network gave rise to path dependence in the location of economic activity. However, several other variables (i.e. historical population density, distance to San Juan, elevation, and suitability for sugarcane cultivation) are stronger predictors of modern

development than distance to historical roads; roads influenced the precise location of economic activity, but geography shaped the general pattern of development. My findings suggest that temporary advantages can have small but persistent effects on city size. Many other papers covering various settings have likewise shown that temporary advantages can have persistent effects on city size (Berger and Enflo, 2017; Bleakley, 2007; Hanlon, 2017; Jedwab and Moradi, 2016; Redding et al., 2011), including colonial infrastructure investments (Bertazzini, forthcoming; Buckwalter, 2019; Graff, 2019; Huillery, 2009; Jedwab and Moradi, 2016; Jedwab et al., 2015; Okoye et al., 2019). Investments in infrastructure intended to disperse development may therefore be ineffective (Banerjee et al., 2020; Bosker et al., 2018).

### 3.2 Transportation in the Late 19th and Early 20th Centuries

For most of modern history, the only important land route in Puerto Rico was a cart road around the edge of the island, though it was incomplete and poorly maintained. Most other routes were essentially natural paths, with horse trails cutting through the mountainous interior. The Spanish government approved plans in 1860 and 1868 for highway construction to facilitate communication between towns and the movement of agricultural products to port. Highways were to be macadamized, or constructed to prevent roads from being washed out in the tropical storms that frequently rendered trails impassable (Martín Ramos, 2005).

These plans were never fully carried out. There were no serious studies of the transportation needs of the various regions, leading to technical and budgetary difficulties. The Spanish governor acknowledged in 1868 that the “almost absolute lack of roads” represented the most powerful impediment to the development of agriculture and commerce, yet Puerto Rico was never able to attract the attention of Madrid to properly finance its infrastructure (Martín Ramos, 2005, p. 46). The system was to span 1,112 km., but only one-fourth was completed by 1896. Construction and maintenance were concentrated in the west, partially because it was flatter and had better natural ports, and the interior was neglected. The only major accomplishment over this period was the *carretera central* (central road), or *carretera militar* (military road), completed around 1886

and stretching 134 kilometers from the capital of San Juan in the north to the commercial hub of Ponce in the south. Other sections of the highways were scattered across the island, incomplete, and detached from one another (Governor of Porto Rico, 1901). There was also a partly finished, single-track, narrow-gauge railroad along several parts of the coastline.

The deficiencies of the transport network through the end of Spanish rule restricted the movement of agricultural products, especially from the interior to the coast. Travel between the east and west was inaccessible to vehicles, and the coastal steamer providing cabotage only circled the island twice monthly. Most products were carried to port on pack animals, at least until they could be loaded on to carriages on flat land. Narrow paths in the mountains required long detours, while rock slides, pot holes, and flooding continued to cut off traffic. Heavy rainfall and a mountainous terrain periodically rendered intercommunication between cities “impossible for weeks at a time” (Governor of Porto Rico, 1901, p. 73).

After the US annexation of Puerto Rico in 1898, American officials acknowledged the dire need for roads. The War Department reported that there were “few roads or ways of communication which are worthy of mention...” (1898, p. 16), and the first annual report of the governor insisted that it was an “imperative necessity to devote every dollar which can be spared from the surplus revenue to the construction of permanent roads” (1901, p. 73). Good roads were deemed “fundamental to the well-being and progress” of Puerto Rico, providing access to schools and facilitating commerce, among other benefits (Carroll, 1899, p. 40). From the outset, officials emphasized the urgency of providing “good roads from seaboard towns into the interior and of substantial trails branching out into the less populous communities” to develop areas lacking transport facilities (Governor of Porto Rico, 1901, p. 427). The construction of roads to benefit the local economy conflicts with claims that public investments in Puerto Rico were intended to narrowly benefit American sugar corporations (e.g., Picó, 1983) or to project military power (Jackson, 2020).

The insular government embarked on an effort to connect all cities and towns with roads. Starting in 1902, road development was guided by seven district boards which were to ensure that

rural areas received a more just share of municipal expenditures but were insulated from parochial concerns by giving municipalities practically no voice. The boards were granted extensive powers but were subject to the authority of the commissioner of the interior and the superintendent of public works, who would coordinate a general plan for all roads and highways (Willoughby, 1905). By the end of 1904, the length of American roads had already surpassed that of Spanish roads. A new route connected Arecibo in the north to Ponce in the south, cutting through the heart of the coffee region and opening up land which previously lacked direct communication with the seacoasts (Governor of Porto Rico, 1904). Many new roads were surfaced to permit heavy traffic of automobiles using modern methods of construction. Transportation costs were estimated to have fallen between 50 and 70 percent in the first decade (Dept. of Commerce and Labor, 1907). Within two decades, each city could be reached by automobile within a few hours from San Juan (Domenech, 1919).

Throughout the 1920s, annual reports of the governor highlighted the ongoing need for roads to “develop the material, social, and intellectual welfare of the people of the interior” (Governor of Porto Rico, 1923, p. 32). In 1916, a plan was enacted to extend highways to connect every important city and town and reach all parts of the interior (Governor of Porto Rico, 1924). By 1925, highways reached all interior towns but did not yet connect them (Governor of Porto Rico, 1926). Even in the mid-1930s reports lamented that roads in the interior were inadequate, but the problem was apparently minor enough to stop appearing in official reports thereafter. Reflecting on achievements in road-building, the 1935 Annual Report pronounced the “first-class system of roads” the “best investment made by the government of Puerto Rico during the 37 years of its existence” and the “principal factor in the development of the island” (Governor of Puerto Rico, 1935, p. 38). According to Diffie (1931, p. 204), “There are few countries which can match the road building achievements of the Porto Ricans in the last generation. . . in regions which would defy a mountain goat.”

The total length of highways increased from 275 km. in 1899 to 2,394 km. in 1940 (Descartes, 1946). Improvements and expansions continued thereafter, but most areas of the island were

connected to the network before World War II.

Railroads also expanded under US rule and eventually reached most coastal towns. The total length of railways increased from 248 kilometers in 1898 to 408 in 1925 and 624 in 1940 (Descartes, 1946). All railroads were privately owned, largely by sugar growers, and survived principally to carry sugar, which constituted more than 86 percent of their freight (Clark, 1930). In 1940, trains carried half of all sugarcane and a fifth of all sugar with only a minimum amount of passenger traffic (de Jesús Toro, 1982). The railroads collapsed shortly after World War II in the face of competition from trucks.

### 3.3 Transportation and Development Data

I constructed a novel dataset by digitizing maps of the Puerto Rican transport network for the years 1899, 1920, and 1940. The maps used were from George F. Cram and the US War Department for 1899, Rand McNally for 1920, and the Puerto Rico Bureau of Public Works for 1940. Several descriptive sources were consulted to classify road quality and verify the accuracy of maps, including the War Department (1898), Governor of Puerto Rico (1901), Carroll (1899), Robinson (1899), and Wilson (2011), as well as a map of road and railroad construction from the Governor of Puerto Rico (1942). I harmonize the locations of infrastructure in 1899 and 1920 with the network in 1940, assuming that minor differences in locations over time are inaccuracies and that the latest map is most accurate.

Using these maps, I conduct analysis at two levels: 2.5 arc-minute grid cells (approximately 4.5×4.5 km.) and municipalities. Data for the grid cells are summarized in Table 3.1 and described in detail in the data appendix. The dependent variables are population density in 1899, 1920, 1940, or 2010; urban extent in 2010; and light density in 2012. Following the literature, these measures proxy for economic activity (Donaldson and Storeygard, 2016). At this level of analysis the primary explanatory variable of interest is (log) distance to the road network from a cell's geographic centroid in 1899, 1920, and 1940, respectively.

I define the primary transport network as consisting of highways, cart roads, and railways, and

Table 3.1: Descriptive Statistics: Grid Cells

	N	Mean	Std. Dev.	Min.	Max.
Distance to roads, 1899 (km.)	496	4.06	3.61	0.01	15.78
Distance to roads < 4 km. in 1899	496	0.60	0.49	0.00	1.00
Distance < 4 km. in 1920, > 4 km. in 1899	496	0.15	0.36	0.00	1.00
Distance < 4 km. in 1940, > 4 km. in 1920	496	0.16	0.37	0.00	1.00
Population density (per sq. km.), 1899	493	122.13	395.90	16.68	8649.07
Population density (per sq. km.), 1920	494	174.78	616.07	24.38	13227.48
Population density (per sq. km.), 1940	494	255.02	805.13	17.02	13494.58
Population density (per sq. km.), 2010	494	445.76	590.03	14.37	5571.49
Urban extent, 2010 (ln)	483	2.18	1.69	-9.34	4.60
Light density, 2012 (ln)	484	3.25	0.67	1.61	4.14
Mean elevation (m.)	496	204.65	212.16	0.03	886.66
Distance to San Juan (km.)	496	60.71	33.73	0.78	130.97
Distance to the coast (km.)	480	9.15	7.11	1.00	26.11
Distance to ports (km.)	496	9.62	0.64	5.43	10.53
Coffee suitability (FAO)	482	3.18	1.26	0.10	7.01
Sugar suitability (FAO)	482	4.40	1.52	1.08	8.67
Tobacco suitability (FAO)	482	2.33	0.80	0.05	4.80
Area (sq. km.)	496	17.62	5.68	0.01	20.35

Notes: See data appendix for descriptions of data. Many variables are expressed in raw form in this table, but all variables are log-transformed for the regressions. All regressions include the same observations by excluding observations with missing values for any variable in the table.

only exclude horse trails, which were reportedly “passable only by the sure-footed native ponies” (Wilson, 2011, p. 235). This classification of roads is historically grounded, as the military highway was universally acclaimed, and the cart road around the island was considered the principal road of the island in the 19th century (Martín Ramos, 2005).<sup>2</sup> Although this paper focuses on roads, railways are included in the network because they were one of the fastest means of transportation on the island. The results are not dependent on the inclusion of railways, however, as major roads often paralleled railways, and railways always comprised only a tiny fraction of the network.

Data at the municipal level are summarized in Table 3.2 and described in detail in the data appendix. All current municipalities are included other than the islands of Culebra and Vieques, or 76 in total. In Puerto Rico, the *municipio* is the principal administrative unit below

<sup>2</sup> Consider, too, the correspondence between my classification and *carreteras de primer orden* (first-order roads) as seen in Figure 1 of Ortiz Colom (2015).

the commonwealth. *Municipio* translates to English as municipality, but *municipios* are US county equivalents. The main dependent variable is population density in 2010, as reported in the 2010 US Census for Puerto Rico. For robustness, I also consider urban extent, light density, per capita income, number of factories during the post-WWII industrialization program, number of corporations, and a socioeconomic development index.<sup>3</sup>

Table 3.2: Descriptive Statistics: Municipalities

	N	Mean	Std. Dev.	Min.	Max.
Market access, 1899 (theta=1)	76	11.91	0.18	11.55	12.47
Population density (per sq. km.), 1899	76	109.04	44.52	31.59	338.14
Population density (per sq. km.), 2010	76	487.92	492.40	66.21	3088.90
Urban extent, 2010	76	2.85	0.79	0.94	4.49
Light density, 2012	76	3.44	0.43	2.29	4.14
Per capita income (USD), 2012-16	76	9889.61	2734.26	5662.97	19701.87
Factories, 1966	76	17.74	27.96	0.00	133.00
Corporations, 2017	76	1420.67	3998.04	47.00	33852.00
Socioeconomic index, 2013	76	54.28	8.63	39.00	82.50
Mean elevation (m.)	76	213.58	170.65	1.64	696.99
Distance to San Juan (km.)	76	55.35	34.67	0.31	124.02
Distance to the coast (km.)	76	18.94	8.59	2.22	37.38
Distance to ports (km.)	76	9.56	6.41	1.29	24.33
Coffee suitability (FAO)	76	3.30	1.02	0.76	5.76
Sugar suitability (FAO)	76	4.40	1.10	1.49	6.68
Tobacco suitability (FAO)	76	2.36	0.59	0.18	3.85
Percentage of land in coffee, 1910	76	8.65	10.33	0.00	44.94
Percentage of land in sugar, 1910	76	8.59	7.85	0.00	36.51
Percentage of land in tobacco, 1910	76	1.13	1.72	0.00	9.32
Sugar (0/1)	76	0.64	0.48	0.00	1.00
Coffee (0/1)	76	0.26	0.44	0.00	1.00
Tobacco (0/1)	76	0.09	0.29	0.00	1.00
Literacy (%), ages 10+, 1899	76	17.83	6.58	8.55	50.24
Agricultural labor (%), 1899	76	68.36	12.01	15.20	86.00
Area (sq. km.)	76	114.66	58.12	13.23	328.53

Notes: See data appendix for descriptions of data. Many variables are expressed in raw form in this table, but all variables are log-transformed for the regressions. The municipalities of Culebra and Vieques are not included in the analysis and do not appear in this table.

<sup>3</sup> The Index of Socioeconomic Development is based on per capita income, median family income, families below the poverty level, unemployment rate, population education level, and illiteracy. The index is modeled on the United Nations Human Development Index. For more information, see the data appendix.

For analysis at the municipal level, I generate an estimate of market access inspired by Harris (1954),  $\sum_{j \neq i} \frac{N_{j,t}}{d_{ij,t}^\theta}$ , where  $N_{j,t}$  is the population of municipality  $j$  in year  $t$ ,  $d_{ij,t}$  is the travel time between municipalities  $i$  and  $j$ , and  $\theta$  is the distance decay parameter. In my baseline specification  $\theta = 1$  (e.g., Berger and Enflo, 2017; Harris, 1954; Klein and Crafts, 2012), but I consider the sensitivity of the results to other values proposed in the literature. Following the standard approach in the literature, I exclude each town’s own population in the calculation. Travel time is estimated using historical transport networks modeled in ArcGIS, and speeds for the various forms of communication are set using historical sources. I do not allow for coastwise shipping because the historical literature provides no indication that cabotage services ever constituted a significant proportion of internal commerce.<sup>4</sup> The logic of the market access variable is that it is advantageous to locate in a municipality which has better opportunities to trade with other municipalities, thereby generating increasing returns to city size. Thus, a municipality’s market access will improve as its travel times to other municipalities—especially large municipalities—fall.

I use market access instead of distance to the road network at this level of analysis for several reasons. First, using an alternative measure serves as a robustness check. Second, market access may be a more appropriate measure than distance to the road network if there is measurement error in the digitized transport network, since the exact location of any given road constitutes only a small fraction of market access. Third, the distance from the centroid to the road network has a less straightforward interpretation at the municipal level because municipalities are much larger than grid cells and are of many different shapes and sizes. Market access incorporates intra-municipality travel, but municipality size has a less direct influence on the measure.

---

<sup>4</sup> Contemporary reports rarely mention maritime trade between Puerto Rican ports prior to US annexation. Martín Ramos (2005, p. 230) claims that “practically one” cabotage company operated throughout the late 19th century and made trips around the island just twice monthly. Smaller boats made trips as well, but natural disadvantages, especially the scarcity of natural ports and the exposure to heavy winds on the north coast, made embarkation and disembarkation difficult and costly. Consider that in 1874, the total volume of cabotage amounted to less than one-tenth of the volume of all imports and exports (de Tejada Morales, 1877). After US annexation, Clark (1930, p. 448) says that coastwise shipping was “relatively unimportant” since American and foreign ships made stops at many island ports.

### 3.4 Roads and Regional Inequality

Roads built under early American rule in Puerto Rico were intended to development the island in general and the mountainous interior in specific. The Spanish transport network favored the largest cities and towns, most of which were on the coast. By contrast, US roads followed a plan to carry out communication from the “interior to the sea” (Governor of Porto Rico, 1902, p. 38). According to Davis (1900, p. 373), “All the systematic work planned or in progress. . . [extends] directly into the interior.”

Since investments targeted remote areas of the island, we might expect a greater concentration of economic activity—and, hence, population—away from areas already connected to the transport network. To test whether this occurred, I run the following regression separately for the years 1899, 1920 and 1940:

$$\ln(\textit{PopulationDensity}_{it}) = \alpha_t + \beta_t \ln(\textit{DistanceToRoads}_{i,1899}) + \varepsilon_{it} \quad (3.1)$$

where  $\textit{PopulationDensity}_{it}$  is the population density of cell  $i$  in year  $t$  and  $\textit{DistanceToRoads}_{i,1899}$  is the distance to the road network (i.e. highway, cart road, or railroad) in 1899. Observations are weighted by land area since some cells on the coast are small. If population growth was higher farther away from Spanish roads, then  $\beta_{1920} > \beta_{1899}$ ; if such differential growth continued beyond 1920, then  $\beta_{1940} > \beta_{1920}$ .

The results appear in Table 3.3 and reveal that population growth after 1899 favored areas closer to Spanish roads. This result holds when excluding areas within 10 km. of the network nodes. In 1899, excluding the areas near nodes, a one percent increase in distance from the transport network corresponded to a 0.05% percent decrease in population density, and R-squared is 0.01. In 1920 and 1940, the relationship between population density and distance to the 1899 network more than doubles and triples in magnitude, respectively, and R-squared increases to 0.09 and 0.16. In other words, the Spanish transport network does not strongly predict population density away from the network nodes in 1899, but population increasingly clustered near the historical network—modernized under early US rule—in the early 20th century despite the rapid construction

Table 3.3: Population Density in Year  $t$  Relative to Distance to Roads in 1899

<i>Dependent variable:</i>	Pop. density, 1899		Pop. density, 1920		Pop. density, 1940	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to roads, 1899	-0.11*** (0.02)	-0.05** (0.02)	-0.18*** (0.02)	-0.12*** (0.02)	-0.25*** (0.03)	-0.18*** (0.02)
Excluded: $\leq 10$ km. from nodes	N	Y	N	Y	N	Y
Observations	470	358	470	358	470	358
R-squared	0.07	0.01	0.18	0.09	0.26	0.16

Notes: The results are drawn from Equation 3.1, a bivariate regression of population density in year  $t$  on distance to the primary transport network in 1899. All variables are expressed as natural logs. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

of roads elsewhere.<sup>5</sup>

Another way to consider this issue is to regress modern development variables on a set of indicator variables equal to one if a cell gained access (< 4 km.) to the transport network before 1899, between 1899 and 1920, or between 1920 and 1940, respectively:

$$\ln(y_i) = \alpha + \beta_1 1[\text{Access}_{1899,i}] + \beta_2 1[\text{Access}_{1900-20,i}] + \beta_3 1[\text{Access}_{1921-40,i}] + \varepsilon_i \quad (3.2)$$

where  $y_i$  is population density, nighttime light density, or urban extent, and the reference group includes cells where the distance to the main transport network was greater than 4 km. in 1940.<sup>6</sup>

The results in Table 3.4 corroborate the finding that development after 1899 favored areas near the historical road network. In 1899, the difference in population density between areas near roads and those that would gain access to roads by 1920 is statistically insignificant ( $p = 0.50$ ). In 2010, the coefficient for roads in 1899 is twice as large as that for roads built between 1899 and 1920, which in turn is more than twice as large as the coefficient on roads built between 1920 and 1940. The results for nighttime light density and urban extent likewise indicate that modern development is greatest in areas near the locations of Spanish transport infrastructure. Figures 3.1, 3.2, and 3.3 illustrate these results visually: many of the most densely populated areas of the island in 1899 fell in the western, interior coffee region which was largely void of major roads, but modern development strongly corresponds to the location of historical roads.

The Spanish transport network was incomplete and generally of low quality but nonetheless would have provided an important initial advantage in the early 20th century. The military highway and some railways were already in operation, and the coastal road allowed for freight transportation by cart. Also, existing cart roads were often upgraded before new roads were constructed, and trails were converted into roads of modern construction (Hull, 1915).

These results do not rule out that areas in the interior would have been worse off without

---

<sup>5</sup> The weak correlation between the location of Spanish roads and population density in 1899 cannot be explained by the fact that the construction of macadamized highways occurred mostly in the last two decades of Spanish rule (see Governor of Puerto Rico, 1919, p. 421). The coastal road had existed for centuries, and some segments of highways had existed in more primitive forms much earlier. For instance, the San Juan-Caguas highway was completed in 1857 (Pumarada O'Neill, 1995).

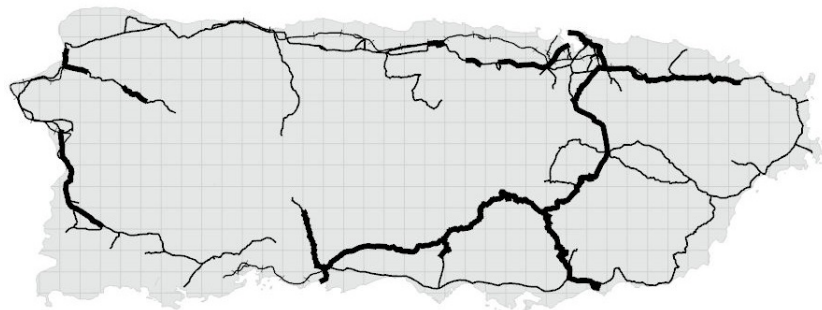
<sup>6</sup> See Figures B.2 and B.3 for maps of the transport network in 1920 and 1940, respectively.

Table 3.4: Modern Development Outcomes and Access to the Transport Network in 1899, 1920, 1940, and 1940

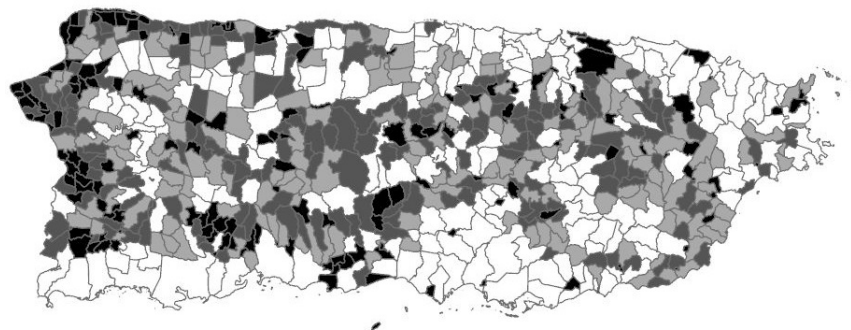
<i>Dependent variable:</i>	Pop. density, 1899		Pop. density, 2010		Light density, 2012		Urban extent, 2010	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1[Distance to roads < 4 km. in 1899]	0.25*** (0.08)	0.15* (0.08)	1.31*** (0.12)	1.15*** (0.11)	1.00*** (0.08)	0.94*** (0.08)	2.44*** (0.23)	2.35*** (0.22)
1[Distance to roads < 4 km. in 1920, > 4 km. in 1899]	0.22*** (0.09)	0.19** (0.08)	0.56*** (0.14)	0.47*** (0.13)	0.37*** (0.10)	0.31*** (0.10)	0.99*** (0.29)	0.91*** (0.29)
1[Distance to roads < 4 km. in 1940, > 4 km. in 1920]	-0.05 (0.09)	-0.08 (0.09)	0.15 (0.14)	0.17 (0.13)	0.16 (0.10)	0.18* (0.10)	0.33 (0.32)	0.42 (0.32)
Excluded: ≤ 10 km. from nodes	N	Y	N	Y	N	Y	N	Y
Observations	470	358	470	358	470	358	470	358
R-squared	0.05	0.05	0.30	0.31	0.37	0.36	0.32	0.30

Notes: The results are drawn from Equation 3.2. The dependent variables are expressed as natural logs. The reference group includes cells that where the distance to the transport network in 1940 was greater than 4 km. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Figure 3.1: Transport Network in 1899

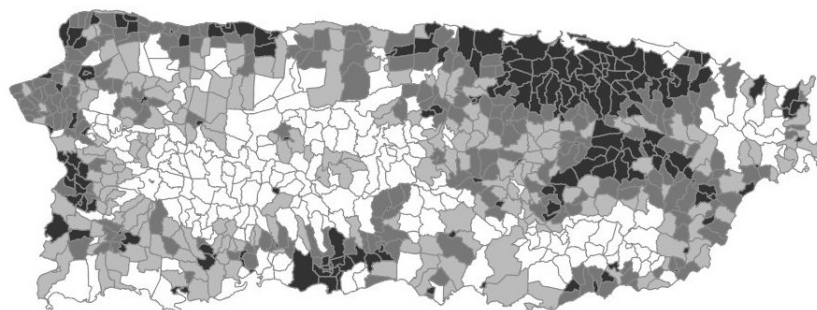


Notes: The transport network is based on maps published by George F. Cram and the US War Department for 1899. The thick black lines are highways; the thin black lines are cart roads; and the grey tracked lines are railways. The grey squares are grid cells used as the unit of analysis.

Figure 3.2: Population Density by *Barrio* in 1899

Notes: *Barrios*, or districts, are divided into quartiles by population density in 1899. Darker shades correspond to higher population density.

Source: Davis (1900)

Figure 3.3: Population Density by *Barrio* in 2010

Notes: *Barrios*, or districts, are divided into quartiles by population density in 2010. Darker shades correspond to higher population density.

Source: US Census Bureau

roads, but new roads failed to pull the spatial distribution of economic activity towards those areas. Instead, areas with initial transportation advantages attracted increasingly greater shares of population even as roads spread throughout the interior. New roads were intended to promote development in the coffee region but appeared instead to facilitate outmigration (Jackson, 2020). Evidence that areas of Puerto Rico previously lacking access to the transport network did not benefit from new roads is consistent with recent research on China and casts doubt on the usefulness of transportation investments in redressing regional inequality (Banerjee et al., 2020; Baum-Snow et al., 2020; Bosker et al., 2018).

### 3.5 The Effect of Roads on Local Development

#### 3.5.1 Grid Data

The previous section demonstrated that modern development outcomes reflect the location of transport infrastructure built prior to the 20th century. In this section, I formally test whether historical access to the transport network affected economic outcomes. I start by regressing development outcomes on distance to the transport network in 1899 and a vector of controls intended to address concerns that historical roads were built in areas more suitable for construction and economic development or where economic activity already existed:

$$\ln(y_i) = \alpha + \beta_1 \ln(\text{DistanceToRoads}_{1899,i}) + x_i' \gamma + \varepsilon_i \quad (3.3)$$

The controls are population density from the first US census in 1899, distance to San Juan, distance to the nearest port, elevation, distance from the coast, and land suitability for the three main crops (sugarcane, coffee, and tobacco).

The results appear in Table 3.5. Distance to roads in 1899 is negatively and statistically significantly related to development outcomes in subsequent years after controlling for observable factors that could determine the location of roads and promote economic development. A one percent increase in distance to roads in 1899 corresponds to 0.05% lower population density in 1920, 0.07% in 1940, and 0.11% in 2010. Modern nighttime light density and urban extent are also

Table 3.5: Persistent Effects of Historical Roads on Economic Outcomes—OLS Results

<i>Dependent variable:</i>	Pop. density, 1920		Pop. density, 1940		Pop. density, 2010		Light	Urban
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance to roads, 1899	-0.05*** (0.01)	-0.05*** (0.01)	-0.08*** (0.02)	-0.07*** (0.02)	-0.07*** (0.03)	-0.11*** (0.03)	-0.12*** (0.02)	-0.38*** (0.07)
Population density, 1899	[0.01]	[0.01]	[0.02]	[0.02]	[0.03]	[0.03]	[0.02]	[0.06]
	0.87*** (0.03)	0.84*** (0.04)	0.89*** (0.05)	0.83*** (0.05)	0.66*** (0.05)	0.67*** (0.05)	0.40*** (0.05)	0.86*** (0.17)
Excluded: < 10 km. from nodes	N	Y	N	Y	N	Y	Y	Y
Observations	470	358	470	358	470	358	358	358
R-squared	0.82	0.78	0.76	0.69	0.73	0.69	0.64	0.38

Notes: The results are drawn from Equation 3.3. The dependent variables are expressed as natural logs. Control variables include elevation; distance from the coast; land suitability for coffee, sugar cane, and tobacco; distance to San Juan; and distance to the nearest port. Robust standard errors are shown in parentheses. Conley standard errors adjusted with a 15 km. cutoff are shown in curly brackets. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

lower in areas farther away from the historical transport network after conditioning on historical development and geographic characteristics. Figures 3.1 and 3.3 illustrate the evident correlation between the historical transport network and modern development, as many of the most densely populated *barrios* are near the locations of roads that existed in 1899. The magnitude of  $\hat{\beta}_1$  is similar in regressions including and excluding cells within 10 km. of the nodes of the historical transport network. The results are robust to correcting for spatial correlation, as seen with Conley standard errors shown in brackets (Conley, 1999).<sup>7</sup>

Reports of the US government state that Spain built roads in Puerto Rico for military purposes and not for private use or commercial development (Governor of Puerto Rico, 1923), as the needs of inhabitants were allegedly neglected entirely (Wilson, 2011). If such claims are true, the location of roads may be exogenous to contemporaneous economic conditions. However, Martín Ramos (2005) maintains that Spanish officials ostensibly cared about building roads to facilitate commerce. Martín Ramos also mentions that roads tended to be built in areas where geographic characteristics made construction easier, and such characteristics might be more conducive to economic activity. In either of these cases, the estimated coefficient on historical distance to the road network would be negatively biased, meaning the effect of old roads on modern outcomes would be overstated.

To address concerns of endogeneity, I follow the literature in implementing an instrumental variables approach (e.g. Berger and Enflo, 2017; Bertazzini, forthcoming; Bird and Straub, 2020; Jedwab and Moradi, 2016; Michaels, 2008; Zimran, 2020). I construct an instrument by drawing straight lines between historical department capitals and between San Juan, the capital, and Ponce, a commercial hub. I then calculate the distance from the centroid of a cell to the nearest straight line and use this distance as an instrument for distance to the road network. Distance from the straight lines is positively correlated with distance to roads but should be uncorrelated with other

---

<sup>7</sup> Following Kelly (2019), I am conservative in setting the cutoff radius and report the largest Conley standard errors. In Appendix B.4, I report Moran's I statistic indicating spatial autocorrelation in residuals; however, I also show using simulations that my measure of distance to the historical transport network consistently outperforms spatial noise in predicting development outcomes, casting doubt on the interpretation that my main results are spurious.

determinants of economic development. The lines connect points of administrative importance when the network was built. I then compare nonnodal locations under the assumption that the only difference between them is their distance to the lines; the strategy is valid if nodes were not chosen so that the line between them would run through economically important regions (Straub, 2015). This assumption is reasonable given that, with the exception of the nodal cities, population tended to be concentrated in the interior, far away from roads, in the late 19th century (see Figures 3.1 and 3.2).

My empirical strategy is grounded in Spanish colonial policy. Infrastructure was intended, above all, to connect points of political importance. The coastal road, for instance, was needed, in part, to connect the capitals of the several departments (de Tejada Morales, 1877). The telegraph system, as authorized in 1866, was designed for governmental functions rather than economic gain. Private construction of telegraph lines was forbidden, and control over lines was not to be devolved to private parties even when favored by Puerto Rican officials. Spain authorized telegraph lines between San Juan and the chief towns of the seven departments. Construction of the lines began in 1872, probably two years after a short line had been built between San Juan and Caguas (Greely, 1900). Therefore, I connect the capitals of the seven departments with straight lines around much of the coast and connect San Juan to Ponce by way of Caguas (see Figure B.1). Other than San Juan, none of these cities is of modern political importance beyond serving as the county seat.

Results based on this instrument appear in Table 3.6. The  $F$ -statistic from the first stage is greater than 10, confirming that the instrument is a strong predictor of proximity to roads. Distance from roads has a negative effect on modern economic outcomes, as areas near roads in 1899 became and remain more developed. Excluding areas near the network nodes, a one percent increase in distance from the transport network reduces population density by 0.10% in 1920, 0.07% in 1940, and 0.24% in 2010; light density in 2012 by 0.40%; and urban extent in 2010 by 0.99%. These effects are all statistically significant at either the 1- or 5- percent level other than for population density in 1940, when the effect is insignificant.

Taken together, these results suggest that historical access to roads promoted local economic

Table 3.6: Persistent Effects of Historical Roads on Economic Outcomes—2SLS Results

<i>Dependent variable:</i>	Pop. density, 1920		Pop. density, 1940		Pop. density, 2010		Light	Urban
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance to roads, 1899	-0.10* (0.05)	-0.11** (0.04)	-0.12* (0.07)	-0.07 (0.06)	-0.32*** (0.12)	-0.24** (0.10)	-0.40*** (0.09)	-0.99*** (0.29)
Population density, 1899	0.84*** (0.05)	0.79*** (0.05)	0.86*** (0.06)	0.83*** (0.06)	0.48*** (0.10)	0.58*** (0.09)	0.21*** (0.08)	0.44* (0.23)
Excluded: < 10 km. from nodes	N	Y	N	Y	N	Y	Y	Y
F-statistic (first stage)	18.7	22.0	18.7	22.0	18.7	22.0	22.0	22.0
Observations	470	358	470	358	470	358	358	358

Notes: Log distance from straight lines connecting the capitals of the seven departments under Spanish rule, and connecting San Juan and Ponce via Caguas, is used as an instrument for distance to the road network in 1899. The dependent variables are expressed as natural logs. Control variables include elevation; distance from the coast; land suitability for coffee, sugar cane, and tobacco; distance to San Juan; and distance to the nearest port. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

development. However, the insignificance of the effect in 1940 should invite some skepticism and highlights the importance of including data from intervening periods in historical persistence research. In addition, the importance of roads in shaping the spatial distribution of economic activity should be interpreted relative to other factors; notably, several other variables (population density in 1899, distance to San Juan, elevation, and suitability for sugarcane cultivation) are stronger predictors of population density in 2010 than distance to historical roads in the OLS and 2SLS results. Roads influenced the precise location of economic activity, but geography shaped the general pattern of development.

The 2SLS point estimates are about two to three times the magnitude of the corresponding OLS estimates. The difference is similar, though less extreme, than in several other studies using distance to straight lines as an instrument (Berger and Enflo, 2017; Bertazzini, forthcoming; Bird and Straub, 2020; Jedwab and Moradi, 2016; Michaels, 2008; Okoye et al., 2019). One explanation for the larger 2SLS estimates is that the local average treatment effect is large, as areas that gained access to the network by virtue of lying between major centers of population benefited more. In particular, towns along the military highway were lucky to have gained early access to the network, since there are now several highways from San Juan to Ponce, the fastest of which does not follow the military highway for much of the way. Another explanation is that the instrument reduces the attenuation bias of the OLS estimates introduced by classical measurement error.

Previous research investigates the sources of persistence by exploring the extent to which infrastructure persists and is coupled with other amenities. In the case of Puerto Rico, such analysis is not possible: all areas of the island are served by schools, hospitals, post offices, and police thanks to Puerto Rico's economic growth over the last century and its continued political affiliation with the United States. Likewise, all areas are connected to the transport network, making placebo tests using proposed roads infeasible.

### 3.5.2 Municipal Data

I turn to municipal-level data using the following specification:

$$\ln(y_m) = \alpha + \beta_1 \ln(\text{MarketAccess}_{1899,m}) + \gamma'_m + \varepsilon_m \quad (3.4)$$

where  $\text{MarketAccess}_m$  is the market access of municipality  $m$  in 1899.  $y_m$  includes the same outcomes as the previous section but also includes per capita income, number of corporations, number of factories, and the Socioeconomic Development Index.  $x_m$  includes the same geographic and development controls as the previous section to account for factors that are related to historical market access and are conducive to economic development.<sup>8</sup>

The results of Equation 3.4 appear in Table 3.7. Column 1 corroborates reports that economic concerns did not dictate the location of Spanish roads: population density is not significantly correlated with market access in 1899, and the point estimate is nearly zero. Columns 2-8 indicate that market access in 1899 is positively and significantly associated with modern economic outcomes. A one percent increase in historical market access results in 1.2% higher population density, 0.4% higher per capita income, and 1.8% more corporations. Municipalities which historically had higher market access also have higher light density, enjoy a higher socioeconomic index rating, and were home to more factories during the post-WWII industrialization.

The positive correlation between historical market access and modern economic outcomes is not driven by outliers. The results are robust to the inclusion of several alternative sets of control variables, as demonstrated in Tables B.1-B.3 of the appendix. These tables, respectively, control for land use in 1910, land suitability, literacy rate, and agricultural employment share. Lastly, Table B.4 shows that the results are robust to replacing  $\theta = 1$  with a wide range of values from the literature.

Market access in 1899 is not significantly correlated with population density in that year. Still, it is possible that historical market access is correlated with unobserved factors that facilitate

---

<sup>8</sup> I use crop dummies in place of FAO land suitability variables in my preferred specification at the municipal level because the FAO measures do not strongly correspond to historical land use. I present results using the FAO variables in Table B.2.

Table 3.7: Persistent Effects of Historical Market Access on Economic Outcomes—OLS Results

<i>Dependent variables:</i>	Pop. Density, 1899 (1)	Pop. Density, 2010 (2)	Urban Extent, 2010 (3)	Light Density, 2012 (4)
Market access, 1899	0.05 (0.23)	1.21*** (0.23)	1.90*** (0.32)	0.84*** (0.15)
Pop. density, 1899		0.64*** (0.11)	0.52*** (0.16)	0.26*** (0.08)
R-squared	0.13	0.79	0.70	0.70

<i>Dependent variables:</i>	Income, 2012-16 (5)	Factories, 1966 (6)	Corporations, 2017 (7)	Socioeconomic Index, 2013 (8)
Market access, 1899	0.39*** (0.12)	1.69** (0.65)	1.78*** (0.52)	0.23** (0.11)
Pop. density, 1899	0.09 (0.06)	0.63** (0.27)	0.78*** (0.27)	-0.00 (0.05)
R-squared	0.59	0.44	0.54	0.27

Notes: All regressions include 76 observations (municipalities). All regressions include the following controls: distance to San Juan, distance to the nearest port, elevation, distance to the coast, and crop dummies. All variables other than indicator variables are expressed as natural logs. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

economic development, which would positively bias the OLS estimate. To address such concerns, I follow the literature in implementing an instrumental variables approach (Jaworski and Kitchens, 2019; Jedwab and Storeygard, 2018). The first instrument is the average travel time for municipality  $m$  to all other municipalities in other districts. Under Spanish rule, Puerto Rico was divided into seven departments. By focusing on travel times to municipalities in other departments, I aim to capture variation attributable to roads not nearby and thus less likely to be driven by local productivity. Also, travel times do not incorporate market size, which is more clearly endogenous to local productivity than are the location of public roads. Since Puerto Rico is small, it is infeasible to limit the instrument to cities, e.g., 50 or 100 km. away, as is done in previous research. Hence, the second instrument is the maximum travel time between municipality  $m$  and all other municipalities. This instrument focuses more narrowly on effective distance to faraway municipalities.

I report the 2SLS results based on average travel time in Table 3.8. The F-statistic from the first stage is high, demonstrating that the maximum travel time is a sufficiently strong instrument for market access. Once again, market access in 1899 is shown to have a large, positive, statistically significant effect on all modern economic outcomes other than the socioeconomic index. Results based on the maximum travel time appear in Table B.5 of Appendix B.2 and indicate that historical market access has a positive effect on modern population density, urban extent, and light density, as well as the number of factories and corporations. For both instruments, the 2SLS point estimates fall within the 95 percent confidence intervals of the corresponding OLS estimates in Table 3.7, and there is no consistent pattern in relative magnitudes between the OLS and 2SLS estimates.

When did market access begin to affect economic outcomes? To answer this question, I use the following generalized difference-in-differences model:

$$\ln(\text{Population}_{mt}) = d_t + \delta_m + \sum_{t=1860}^{2010} \beta_t \ln(\text{MarketAccess}_{1899,m}) \cdot d_t + \varepsilon_{mt} \quad (3.5)$$

with 1846 as the reference year.  $\beta_t$  represents how the relationship between market access and municipal population in year  $t$  differs relative to 1846. The pattern of results is generally unchanged by using market access for a different year or by setting all city populations equal to one when

Table 3.8: Persistent Effects of Historical Market Access on Economic Outcomes—2SLS Results (Mean Travel Time)

<i>Dependent variables:</i>	Pop. Density, 1899	Pop. Density, 2010	Urban Extent, 2010	Light Density, 2012
	(1)	(2)	(3)	(4)
Market access, 1899	-0.08 (0.28)	1.09*** (0.28)	1.96*** (0.40)	0.84*** (0.19)
Pop. density, 1899		0.64*** (0.10)	0.52*** (0.15)	0.26*** (0.08)
F-statistic (first stage)	209.8	201.3	201.3	201.3

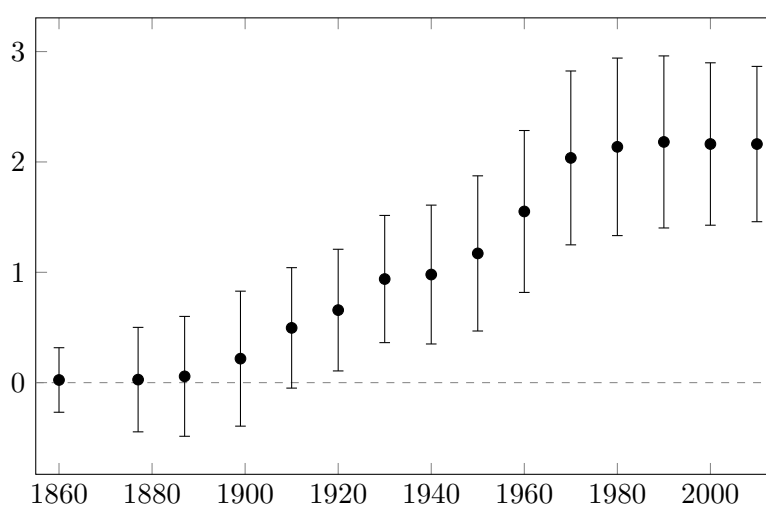
<i>Dependent variables:</i>	Income, 2012-16	Factories, 1966	Corporations, 2017	Socioeconomic Index, 2013
	(5)	(6)	(7)	(8)
Market access, 1899	0.25** (0.12)	1.96** (0.77)	1.69** (0.68)	0.03 (0.11)
Pop. density, 1899	0.09* (0.05)	0.62** (0.26)	0.78*** (0.25)	-0.00 (0.05)
F-statistic (first stage)	201.3	201.3	201.3	201.3

Notes: The mean estimated travel time in 1899 to all municipalities in other (defunct) departments is the instrument for market access in 1899. All regressions include 76 observations (municipalities). All regressions include the following controls: distance to San Juan, distance to the nearest port, elevation, distance to the coast, and crop dummies. All variables other than indicator variables are expressed as natural logs. Robust standard errors are shown in parentheses. Significance levels are denoted by \* p < .10, \*\* p < .05, \*\*\* p < .01.

estimating market access.

I plot the estimates of  $\beta_t$  in Figure 3.4. Throughout the 19th century, the relationship between market access and city size does not change. Then, from around 1899 to 1970, there is an increasingly positive relationship. After 1970, the relationship levels off, such that the positive correlation between historical market access and modern city size is largely the product of change that occurred in the first six decades of the 20th century.

Figure 3.4: Market Access and Population Density, 1846-2010



Notes: The regression coefficients and 95 percent confidence intervals come from Equation 3. The coefficients represent the effect of market access on population density relative to the year 1846.

In other words, roads mattered more for local economic development after the turn of the 20th century. During that period, Puerto Rico transformed from a “largely self-sufficing island economy” into one which was “integrally related with the commerce of the rest of the world, particularly... of the United States” (Clark, 1930, p. 403). By the 1910s, the per capita volume of external trade was nearly four times higher than it had been throughout the 19th century (US Tariff Commission, 1943), and Puerto Rico was the largest purchaser of US products in Central and South America other than Argentina (Governor of Puerto Rico, 1910). Such economic changes “swamped the avenues of trade throughout the island...”, generating a “great need” for roads (Governor of Porto Rico, 1907, p. 13).

### 3.6 Conclusion

This paper considers the effect of roads on regional inequality and local economic development by studying a program of road-building covering an entire territory—Puerto Rico. With a few decades, highways covered the entire island, yet spatial inequality began to worsen during this period (see Chapter 4). The population was somewhat uniformly distributed across the island at the turn of the 20th century and is now highly concentrated around San Juan. Roads were supposed to bring prosperity to the western highlands, which lacked access to the transport network, but population growth lagged in that region. During the second half of the 20th century, Puerto Rico experienced the largest rate of reforestation on Earth as the decline of the interior intensified (Rudel et al., 2000). Some of these low-density areas are intersected by major highways (Martinuzzi et al., 2007).

Early access to the transport network had a positive effect on economic development in the short run, and the effects of the advantage persist to the present even after other areas were connected to the network. These results are consistent with the existence of increasing returns to city size and serve as further evidence of diminishing marginal effects of infrastructure investments. However, it is important to consider that other factors, especially topography and distance to San Juan, played a more important role in the general spatial pattern of development in Puerto Rico; it is difficult to imagine a reality in which economic activity does not cluster in the coastal lowlands. Hence, the evidence from this paper suggests a limited role for transport infrastructure to shape the economic distribution of economic activity, although that role may vary by context.<sup>9</sup>

---

<sup>9</sup> Tax policy intended to benefit peripheral regions of the island were likewise ineffective (Guimaraes et al., 1998).

## Chapter 4

### Spatial Population Trends and Economic Development in Puerto Rico, 1765-2010

A version of this chapter is forthcoming in the *European Review of Economic History*.

#### 4.1 Introduction

This paper examines long-run spatial patterns in population growth in Puerto Rico, whose unique economic history offers insight into the development process. The Caribbean colony was neglected by Spain for centuries and relied almost entirely on subsistence agriculture as late as the early 19th century, lagging behind most of the West Indies (Mintz, 1974). Even after the rise of commercial agriculture, Puerto Rico remained one of the less developed areas of Latin America, as gross domestic product per capita declined over the late 19th century (Santamaría García, 2012). Following annexation by the United States in 1898, the island emerged as one of the leading sugar producers in the world, then rapidly industrialized and ranked among the fastest growing economies after World War II (Baumol and Wolff, 1996).

I use newly standardized, municipal-level data to quantify regional patterns of population growth from 1765 to 2010. First, I present a time series of the correlation between historical and modern population densities to assess persistence in the spatial distribution of population. Population density throughout the 18th and 19th centuries is weakly correlated with modern population density. After 1899, the distribution of population across municipalities quickly approached the modern distribution, as the greatest change in relative population density occurred

between 1899 and 1910. The postwar period is visibly another period of change, as the population clustered in San Juan and, increasingly, in nearby municipalities. These results indicate that the turn of the century—not just the postwar period, as suggested in previous research—was a turning point for economic development. Second, I analyze several measures of variation in municipal population density. Variation was high in the 18th century, with the population concentrated on the coast, but fell throughout the 19th century as the population spread across the island. The population was more uniformly distributed across municipalities around the turn of the 20th century than at any other point in recorded history. Thereafter, variation increased as population clustered in San Juan and surrounding municipalities, accelerated after 1940, and plateaued around 1980. Third, population growth was positively correlated with traditional crop production—coffee in the late 19th century, and sugarcane and tobacco in the early 20th century—in the preindustrial era and was negatively correlated with agricultural employment from 1899 to 1970.

One of the major implications of these results is that urbanization began around 1900—decades earlier than generally believed—when variation in municipal population density began to rise and population growth was higher in areas less reliant on agriculture. This result is important for understanding Puerto Rico’s economic trajectory since development is closely intertwined with increasing regional variation in population density as well as a shift of labor out of agriculture (Davis and Weinstein, 2002; Henderson, 2010; Michaels et al., 2012). Past research tends to treat post-WWII industrialization as a sharp break from a stagnant past (e.g., Baer, 1959). Yet, the spatial trends exhibited during the postwar “miracle” were a mere continuation of trends originating at the turn of the century, indicating that Puerto Rico’s impressive socioeconomic transformation in the 20th century began, contrary to most accounts, before World War II.

To formally address urbanization and its origins, I construct a time series for an urbanization index using standardized definitions of metropolitan areas proposed by International Urban Research (1959). The share of the population living in the municipalities that would become the main centers of population by 1950 hovered around 12 percent from 1846 to 1899, then grew continuously such that by midcentury Puerto Rico, at 27 percent, trailed only the Southern Cone

within Latin America and was on par with Cuba and Venezuela. To place Puerto Rican urbanization in a regional context, I also compute the index for Central America and the Caribbean for 1850 to 1950 and find that the timing of urbanization in Puerto Rico is unique. Consistent with Browning (1958) and Boswell (1978), I show that most of these countries did not urbanize until after World War II. The only exceptions are Panama, after the opening of the Panama Canal, and Cuba, which was already highly urbanized in 1900. The conclusion that urbanization in Puerto Rico began around 1900, and that urbanization in many small countries in the Circum-Caribbean lagged by several decades, motivates additional research.

I draw on 11 censuses between 1765 and 1899 and decennial censuses from 1910 forward. Puerto Rico is unique within Latin America in its abundance of census data prior to the 20th century, but previous research has been unable to evaluate long-run spatial trends because municipal boundaries changed over time. I standardize municipal definitions across previously digitized population counts from 1765 and 2010 by aggregating municipalities on the basis of historical groupings in a manner analogous to Horan and Hargis (1995). This new panel dataset allows for an analysis of long-run spatial trends and sheds light on early Latin American economic development.

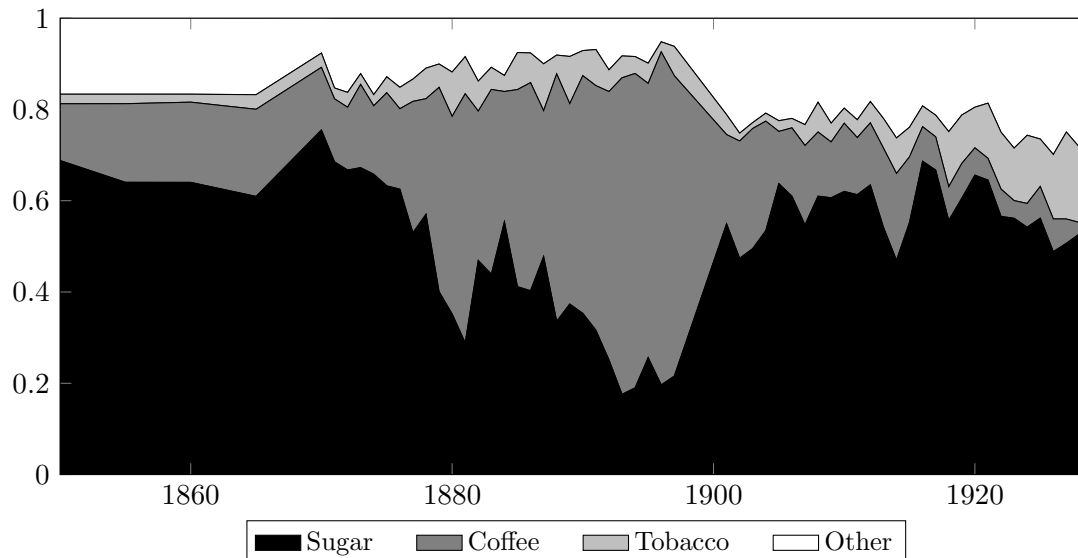
I focus on long-run trends in municipal population growth to comment on economic development for two reasons. First, population data are available over a longer period and are more comparable over time than are other development indicators. A large literature in economic geography uses population growth as a proxy for economic dynamism and population density as a proxy for economic activity (e.g., Beeson et al., 2001; Davis and Weinstein, 2002). Studying Puerto Rico from a long-run perspective is important for understanding when and, ultimately, why its fortunes turned. To my knowledge, this paper marks the first effort to standardize data for comparison across most of Puerto Rican history since the arrival of Europeans—most importantly, between Spanish and US rule. Second, although urbanization per se does not cause development, sustained economic development does not occur without urbanization, and there is a strong, positive cross-country relationship between urbanization and per capita GDP (Henderson, 2010). An analysis of spatial trends, therefore, enriches our understanding of economic development.

## 4.2 Regional Development Throughout History

At the time of the first census in 1765, Puerto Rico had a population of just 44,833, mostly residing on the coast. Madrid valued Puerto Rico as a naval outpost to protect sealanes to other colonies, leading to a policy of neglect (Dietz, 1986). The island received little immigration and relied mostly on subsistence agriculture as late as the early 19th century, lagging behind most of the West Indies (Mintz, 1974). The early 19th century was marked by reforms intended to encourage economic development, including a period of free trade with Spain, the reduction of tariffs and opening of all ports to foreign trade, the adoption of the Spanish commercial code, and the opening of borders to Catholic immigrants. Land grants and tax exemptions attracted immigrants, helping population to more than quadruple between 1815 and 1899. Immigrants arrived especially from Spain and from Latin America fleeing the wars of independence and bringing with them skills and capital. The importation of slaves, mostly to work in coastal sugar fields, increased in the late 18th and early 19th centuries, but slaves never accounted for a large share of the labor force—just five percent of the population at abolition in 1873, the smallest proportion of any Caribbean colony in the 19th century (Dietz, 1986).

The three most important exports historically were coffee, sugar, and tobacco. Sugar was the island's most valuable export for most of the 19th century (see Figure 4.1), prospering on the coastal plains in a time of rising prices after the collapse of the sugar industry in Saint Domingue and the gradual decline of plantation agriculture in the British and French Caribbean (Ayala and Bergad, 2020). The industry stagnated around midcentury, however, as producers failed to modernize production in an era of greater international competition including the rise of European beet sugar. Puerto Rico lacked the capital and financial institutions required to invest in new technology, and technological development was blocked by restrictions on credit and the importation of industrial machinery (Dietz, 1986). Meanwhile, coffee, which required minimal capital investment and could be grown alongside subsistence crops, thrived in the mountains of the sparsely populated western interior, where land was available for settlement as private property or in usufruct (Ayala and

Figure 4.1: Shares of Total Export Value by Product, 1865-1929



Sources: Miller (1922), US Department of Commerce and Labor (1907), and Clark (1930).

Notes: Data are unavailable from 1898 to 1900. Tobacco does not include manufactured products.

Bergad, 2020). In the 1870s, coffee overtook sugar and remained the most important agricultural product through the end of the century. In 1850, coffee and sugar, respectively, represented 12 and 69 percent of the total value of exports; in 1897, coffee made up 66 percent of total export value and sugar just 22 percent. Lastly, tobacco exports remained marginal until the second half of the 19th century, when high-quality tobacco production developed in the eastern central highlands for export to Cuba for cigar production.

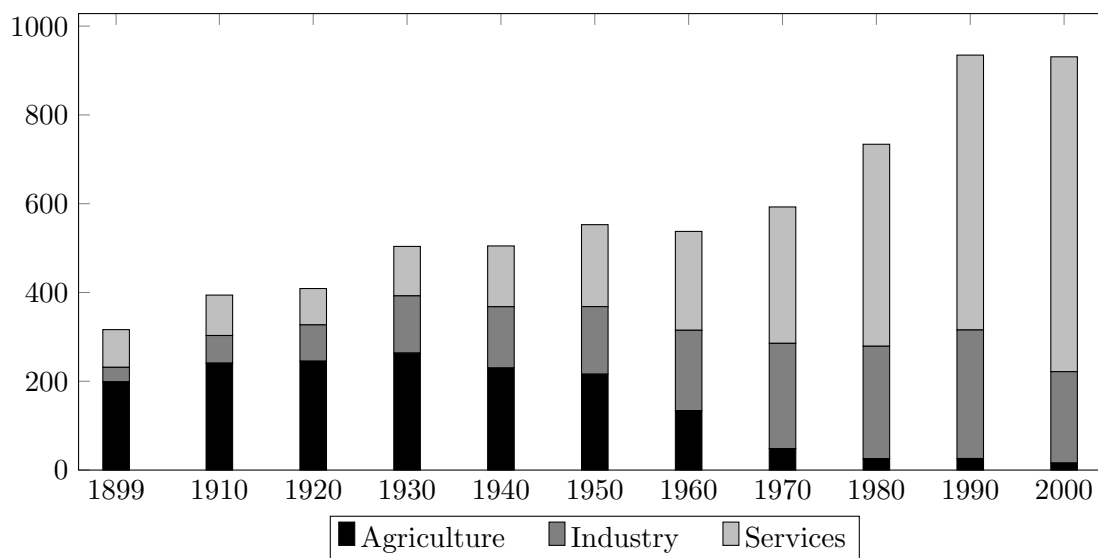
Spain ceded Puerto Rico to the United States at the conclusion of the Spanish-American War in 1898. The US passed the Foraker Act in 1900, establishing a civilian government and a policy of free trade with the US. The production of cane sugar, which enjoyed protection under the US tariff regime, grew from 66,073 tons in 1895 to 866,109 tons in 1930 (Dietz, 1986). Fueled by US corporate investment, sugar estates were transformed into *centrales*, or huge grinding mills which processed the cane from a large surrounding area using modern technology. Tobacco also benefited from tariff protection and prospered especially in the eastern central highlands. US capital quickly took control of tobacco manufacturing, partially mechanizing production. Figure 4.1 understates

the importance of tobacco, since the share of total export value from tobacco does not include manufactured products and manufactured exports exceeded the value of raw exports until the end of World War I. By 1905, tobacco (raw and manufactured) displaced coffee as the second most valuable export, and by 1918 exports of raw tobacco surpassed exports of coffee in value. Tobacco exports declined in the 1930s, from 17.5 percent of total export value in 1931 to just 6.6 percent in 1940, and never recovered. Coffee, which enjoyed no tariff protection, quickly lost its place as the most valuable export, falling to 19 percent of export value in 1914 and less than one percent in 1935. The decline of coffee as a share of exports was due mostly to the explosive growth of sugar and tobacco production, not a reduction in the volume of coffee produced. Coffee exports dropped precipitously and never recovered after hurricanes in 1928 and 1932, but some coffee production for local consumption remained (Perloff, 1950). Lastly, citrus fruit and coconut farming developed in the early 20th century, concentrating in the municipalities on the north coast of the island. Coconut production was never large, but the value of citrus fruit exports exceeded that of coffee exports by the 1920s.

The economy also gradually shifted away from agriculture and towards manufacturing, as shown in Figure 4.2. From 1899 to 1940, employment in agriculture as a percentage of all jobs fell from 62.8 to 45.5 percent, while employment in industry grew from 10.3 to 27.3 percent. The principal manufacturing sectors were sugar and tobacco processing, with mills and factories located in the regions of their raw materials and those producing for local consumption located in the centers of population (Picó, 1974). Other manufacturing sectors expanded as well, including fruit canning and cotton goods. The manufacture of cotton goods was typically carried out at home and mostly by women in major urban areas (Bureau of Labor Statistics, 1943). Manufactured cotton goods grew steadily but represented a tiny fraction of all exports until the 1920s, constituting more than a tenth of total export value by the end of the decade (Clark, 1930).

After World War II, Puerto Rico launched an industrialization program, including a ten-year exemption on insular income and property taxes for new industries, to attract US firms to the island. In the first decade of the tax exemptions, private investment rose 377 percent while real per capita

Figure 4.2: Labor Force by Industry (in thousands), 1899-2000



Source: US Census Bureau.

income rose 59 percent and surpassed the rest of Latin America (Baer, 1959). Industrialization favored established centers of population in general and Greater San Juan in specific. More than one-third of the 2,367 manufacturing enterprises established under the industrialization program and in existence in 1967 were in the municipalities of San Juan, Cataño, Bayamón, Guaynabo, and Carolina (Picó, 1974).

The postwar “miracle” ended in the early 1970s. Income had been converging to the US since the mid-1940s, then growth slowed, convergence reversed, and a prolonged period of stagnation commenced (Lefort, 1997). Net transfers, mostly from the federal government, jumped from ten percent to one quarter of gross national income in the 1970s (Devereux, 2019). The unemployment rate rose from 10.7 percent in 1970 to 23.4 percent in 1983, and the labor force participation rate for men began a sharp decline in the 1970s from which it has never recovered (Department of Labor and Human Resources, 2010). Unemployment was higher in rural areas of the interior than in urban areas, especially those near San Juan (Rivera-Batiz and Santiago, 1996). Manufacturing as a share of total employment has declined since 1978, with services now providing most employment (Dietz, 2003). In the 1990s, tax incentives offered to US firms operating in Puerto Rico were weakened and

ultimately repealed, accelerating the decline of manufacturing (Feliciano and Green, 2017). The tax incentives were finally phased out in 2006, around which time the insular government suffered a budget crisis and the economy entered a deep, long recession.

### 4.3 Data

I standardize municipal definitions across previously digitized population counts from 1765 and 2010 to construct a panel dataset with which I provide an overview of trends in the spatial distribution of population throughout Puerto Rican history. Raw population counts come from censuses of Puerto Rico, conducted by Spain from 1765 to 1897 and by the US from 1899 to 2010. Data for most Spanish censuses and for US censuses through 1970 were transcribed from Vázquez Calzada (1978) and for US censuses from 1980 to 2010 from <http://www.statoids.com/upr.html>. The 1790 and 1800 Spanish censuses, or *padrones*, were downloaded from Scarano and Curtis (2011). In total I draw on 23 censuses, typically separated by ten years but longer between the earliest censuses. Spanish censuses of Puerto Rico were conducted irregularly until 1877, after which they took place every ten years. There are slightly larger gaps in the data between 1800 and 1827 and between 1827 and 1846 because data because the 1815 and 1832 censuses are not easily accessible. US censuses of Puerto Rico are conducted decennially in conjunction with the US census, with the exception of the first census in 1899. A special census was also conducted in 1935 by the Puerto Rico Reconstruction Administration, but I rarely use this census in the interest of consistency. A Pearson correlation coefficient of 0.99 between municipal populations from the last Spanish census in 1897 and the first US census in 1899 provides strong justification for treating the Spanish and US censuses as one continuous time series and all but eliminates the concern that demographic changes between the two periods are the product of systematic differences in measurement between the empires.

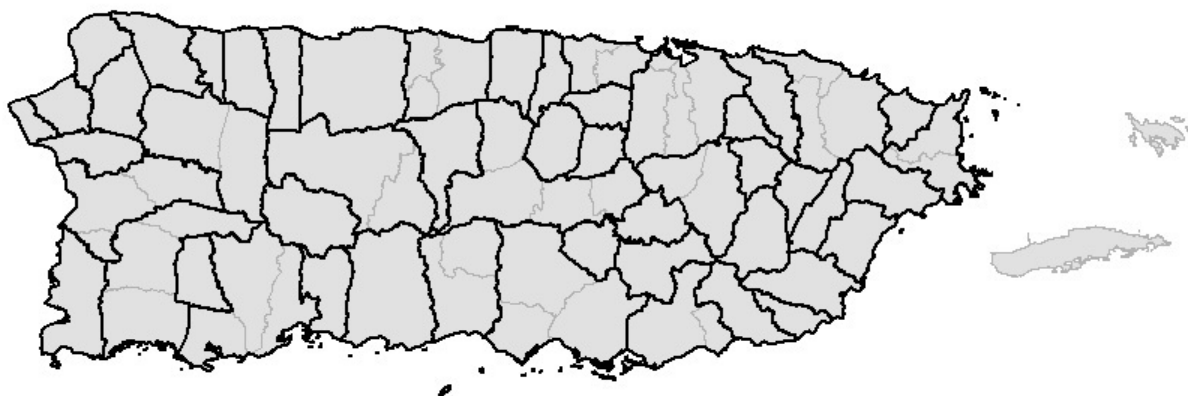
The level of analysis is the *municipio*, which translates to English as “municipality” but more closely resembles a county in US states. In the first Spanish census in 1765 there were 24 municipalities. New municipalities emerged with each Spanish census as population grew and

segments of municipalities became sufficiently populated to break off. Currently there are 78 municipalities, comprising the entirety of the commonwealth. To generate consistent definitions of municipalities over time, I homogenize municipalities based on the municipal definitions from the earliest census used in a given exercise, relying on municipality histories from the online Encyclopedia of Puerto Rico published by the Puerto Rican Endowment for the Humanities (2018). My aggregation strategy groups municipalities, as defined by modern boundaries, into larger units on the basis of historical configurations in a manner analogous to Horan and Hargis (1995). A detailed account of the aggregation is provided in Appendix C.2. A map of municipalities standardized for analysis starting in 1827, and in relation to modern municipalities, is displayed in Figure 4.3.

The advantage to analyzing municipalities is that it allows for precise observation of population movements. In my most aggregated data set, using municipality definitions from 1765, the mean land area is 544 square kilometers with a standard deviation of 303. In the dataset based on the 1827 census, which I rely on in the regression analysis, the average area is just 171 square kilometers with a standard deviation of 115. Additionally, geographical units that span the whole area of the polity under examination remove the possibility that results are unique to the sample selected by the researcher. The main drawback to studying municipalities is that the use of geographically small units in the presence of suburbanization gives an exaggerated impression of a shifting equilibrium. In any case, the municipality is the primary administrative division.

In addition to population data, I incorporate crop production and development data to test for correlates of municipal population growth and net migration rates. These variables are described in detail in the data appendix, and summary statistics are found in Table 4.1.

Finally, the reader may be concerned about nonrandom measurement error in the longitudinal population and area data. In the absence of historical maps with delineated municipal boundaries, I treat borders as unchanging and largely ignore measurement error. However, there is no obvious reason to believe that measurement error for population growth would be systematically correlated with right-hand side variables, and the effect of an unobserved boundary change would last only one

Figure 4.3: Map of Standardized Municipalities (*Municipios*) in 1827

Notes: The black lines represent the borders of municipalities grouped together based on historical municipal definitions (and modern borders) to allow for analysis with consistently defined units starting in 1827. The grey lines represent current boundaries where they do not overlap with municipal borders based on units aggregated for 1827. For more information about on the standardization of municipal boundaries, see Appendix C.2. Vieques and Culebra, the smaller islands to the east of the main island, were not part of Puerto Rico in 1827.

Table 4.1: Descriptive Statistics

	Munis.	Mean	Std. Dev.	Min.	Max.
Percentage literate, ages 10 and up, 1899	69	18.17	7.79	8.55	61.26
Percentage of labor in agriculture, 1899	69	67.83	13.49	1.41	86.00
Percentage of labor in manufacturing, 1899	69	6.58	4.39	1.77	23.55
Percentage of land in coffee, 1910	67	7.76	10.17	0.00	44.94
Percentage of land in sugar, 1910	67	7.83	7.65	0.00	36.51
Percentage of land in tobacco, 1910	67	1.18	1.82	0.00	9.32
Sugar (0/1)	77	0.65	0.48	0.00	1.00
Coffee (0/1)	77	0.26	0.44	0.00	1.00
Tobacco (0/1)	77	0.09	0.29	0.00	1.00
Mean annual rainfall (inches)	75	72.12	18.70	33.05	135.55
Mean altitude (meters)	76	213.59	170.65	1.64	696.99
Mean land gradient (degrees)	75	12.00	4.44	3.22	20.69

Notes: The descriptive statistics are derived from raw data. In the statistical exercises some municipalities are grouped together to reflect historical municipal definitions and municipalities with missing data are assigned values based on those of neighboring municipalities.

period. Since this paper is an analysis of trends, measurement error in population growth would need to be correlated with the right-hand side variables repeatedly over time, and this is unlikely.

## 4.4 Municipal Population Density Over Two and Half Centuries

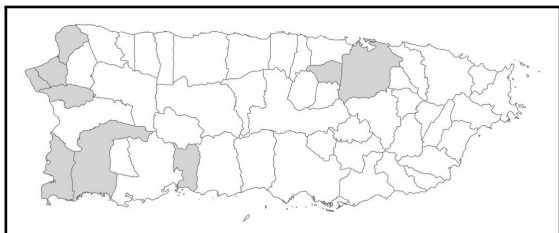
### 4.4.1 Persistence in Municipal Population Density

The first question to address is whether there is persistence in relative municipal population density throughout Puerto Rican history. Were the most densely populated areas of modern Puerto Rico also the most densely populated historically? If not, when do we see changes? Modern Puerto Rico is developed and urbanized, so determining when the distribution of population started to resemble its modern form can provide insight regarding the commencement of urbanization and modernization of the economy.

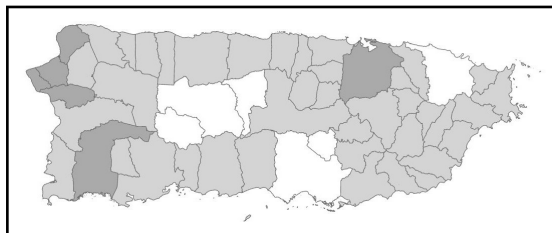
I start by observing heat maps of population density for selected years in Figure 4.4. In 1827, the island was lightly populated, and the west coast and the area near San Juan were the most densely populated. Ponce also had one of the largest populations on the island, but its large area was such that population density for the municipality was not high. By 1860, the population had doubled and spread throughout the island. The population became less concentrated on the west coast, though the most densely populated municipalities by and large did not change. There was a marked movement to the interior of the island—the “internal frontier” (Ayala and Bergad, 2020, p. 46)—but the western interior remained sparsely populated. Over the next 40 years, coffee prospered in the western highlands, and the municipalities in this region ranked among the most densely populated by the end of the century. From 1899 to 1940, the coast in general and San Juan in specific became relatively much more populous. In 1899, the ratio of the populations of San Juan (including Río Piedras) and Ponce was 0.78; in 1940, the ratio had grown to 2.26. Among the other more densely populated municipalities in 1940 were Caguas and nearby municipalities in the eastern interior as well as several port cities, including Ponce, Mayagüez, Aguadilla, Arecibo, and Humacao. After 1940, the maps reflect greater concentration of population in the northeast coast

Figure 4.4: Municipal Population Density (per sq. km.), 1827-2010 (Selected Years)

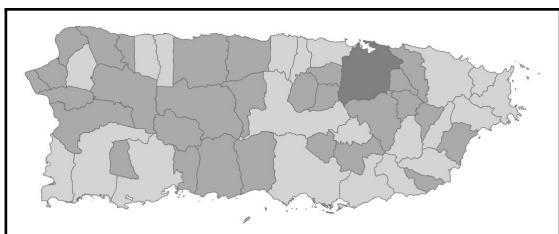
(a) 1827 (Pop.: 280,734)



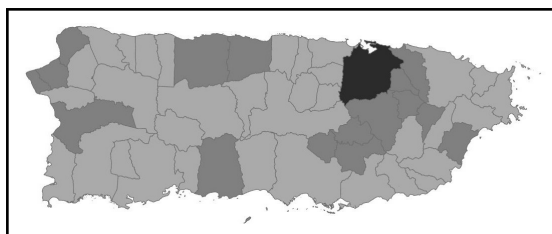
(b) 1860 (Pop.: 583,308)



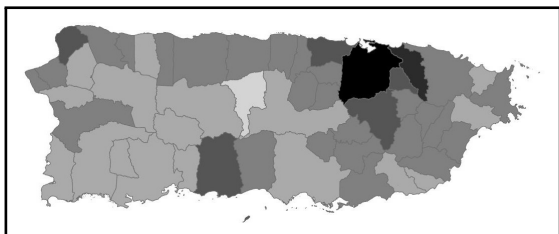
(c) 1899 (Pop.: 953,243)



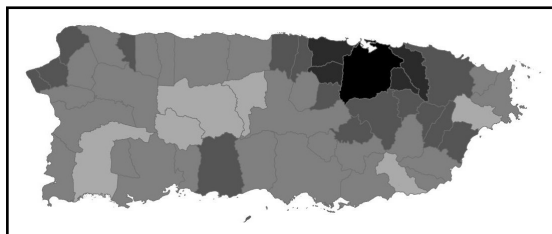
(d) 1940 (Pop.: 1,869,255)



(e) 1970 (Pop.: 2,712,033)



(f) 2010 (Pop.: 3,725,789)



Notes: Municipalities in these maps are grouped into larger units on the basis of historical configurations to allow for analysis with consistently defined units starting in 1827.

Source: Author's calculations based on population censuses.

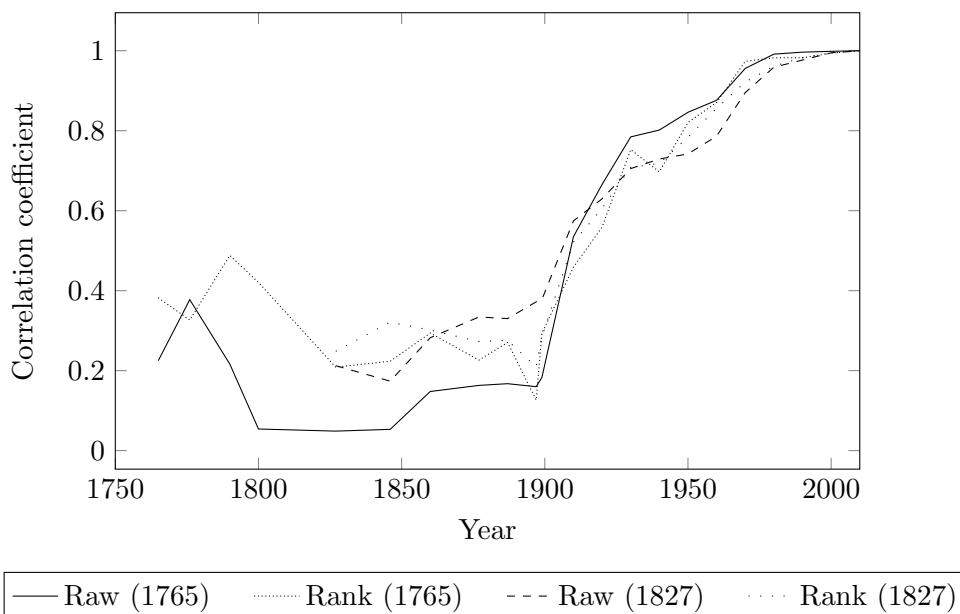
of the island, particularly in the municipalities closest to San Juan. The share of the population residing in the San Juan-Caguas-Guaynabo Metropolitan Statistical Area (MSA), as currently defined, increased from 0.47 in 1899 to 0.67 in 2010; meanwhile, the share of the MSA population residing in San Juan proper (including the former Río Piedras) grew from 0.10 in 1899 all the way to 0.32 in 1960 but has since fallen to 0.16. The western interior has once again become much less populated relative to the rest of the island. Many of the municipalities in that region experienced net population declines between 1940 and 1970, and several (Utuado, Ciales, Adjuntas, and Maricao) experienced net declines between 1940 and 2010. In sum, the spatial distribution of population varied considerably over time.

Following Davis and Weinstein (2002) and others (e.g., Ayuda et al. (2010a,b); Bosker et al. (2008)), I formally assess persistence in population density by estimating and describing temporal patterns in correlation coefficients for historical municipal population density with population density in 2010, displayed in Figure 4.5. Correlation should increase over time, but the analysis helps to identify periods of significant change.

I start by aggregating data to 1765 municipal definitions, leaving 16 units, to be able to use all years of data. Although there were 24 municipalities in the 1765 census, I combine some municipalities to maintain consistent definitions across censuses. Correlation with modern population density is low throughout the 18th and 19th centuries and rises only to 0.16 in the last Spanish census in 1897. Between 1899 and 1910, raw correlation jumps to 0.54 and continues upwards until it has essentially reached the modern distribution in 1970 at the latest with a correlation of 0.96. Rank correlation, which is less influenced by outliers, parallels this trend, as do correlations based on municipal definitions from 1827, which allow data to be disaggregated into 51 observations for each census year. The pattern is essentially unaffected by the level of aggregation.

Decadal changes in municipal population density increased anew with industrialization after World War II. But in some sense the ostensible change after 1950 is a statistical illusion resulting from the small area of municipalities. While this exercise gives the impression that new important

Figure 4.5: Correlation Between Historical Municipal Population Density and Municipal Population Density in 2010



Notes: Municipalities have been aggregated to allow for analysis with consistently defined units beginning in the year in parentheses in the legend. Years on the x-axis represent the census year for which correlation with municipal population density in 2010 has been estimated. See Section 4.4.1 for more details.

Source: Author's calculations based on population censuses.

cities emerged after 1950, the reality is that San Juan overflowed beyond its *de jure* limits. For example, when population density in 2010 is regressed on population density in 1950, the municipalities with the largest residuals by far are all near San Juan (Carolina, Toa Alta, Toa Baja, and Trujillo Alto), and their residuals are all positive. Recall that San Juan in this analysis includes the municipalities of Cataño, Bayamón, and Guaynabo to maintain consistent boundaries over the sample period. When the municipalities near San Juan are excluded, correlation with municipal population density in 2010 is much higher—0.95 instead of 0.74 in 1950, for example, when using 1827 boundaries.

The next question then becomes, to what extent was the spatial distribution of population stable prior to 1899? I perform the same exercise but compare municipal population for a given census year with the corresponding figure for 1897 (not shown). Moving backwards, the spatial

distribution appears similar as far back as 1877, which has a correlation of 0.90. There is then a considerable drop moving back to 1860 as correlation falls to 0.66. This result comports with historical accounts of a reorientation of the economy from sugar to coffee cultivation in the 1870s. Before 1860, correlation is even lower (e.g., 0.28 in 1765).

To summarize, persistence in relative population density is low before the 20th century. Changes in relative population density correspond to some documented economic changes, including the coffee boom in the late 19th century and post-WWII industrialization. But the clearest period of change over the 245 years of data is the beginning of the 20th century. The shifting patterns of population growth in these periods will be discussed in Section 4.4.3.

#### **4.4.2 Variation in Municipal Population Density**

I now turn to an examination of variation in municipal population density over time. The spatial concentration of population is a key feature of economic development. In the early stages of modern growth and industrialization, agglomeration forces lead economic activity to become spatially concentrated (Martinez-Galarraga et al., 2020). Industrialization in other settings—for example, in Spain, Japan, and the United States (Ayuda et al., 2010a; Davis and Weinstein, 2002; Michaels et al., 2012)—coincided with greater concentration of population, since the forces of agglomeration are stronger outside of agriculture. According to theories of increasing returns, advantages to city size can arise from knowledge spillovers, labor-market pooling, or the advantages of proximity for both firms and consumers. Trade is also a key driver of the spatial distribution of population and economic activity, and locations with good access to trade, such as harbors, tend to have more people (Nagy, 2018).

Time series of five measures of variation in municipal population density are displayed in Table 4.2. Higher values of each measure signify greater variation. The first three measures are the ratio of mean to median population density, the ratio of the 90th to 10th percentile population density, and the share of population living in the five most populous municipalities in a given census year. The fourth measure, the Gini coefficient, equals zero in the case of population being

Table 4.2: Variation in Population Density

Year	Population (thousands)	Mean/median muni. pop. density	90th/10th percentile	Share of five largest munis.	Gini coefficient	Rel. var. of log pop. density
1765	45	1.48	10.34	0.57	0.49	2.34
1776	71	1.37	12.21	0.58	0.42	2.04
1790	107	1.61	9.14	0.59	0.43	1.76
1800	156	1.52	7.26	0.58	0.42	1.51
1827	281	1.22	5.29	0.55	0.31	0.93
1846	445	1.14	3.72	0.52	0.23	0.53
1860	583	1.07	2.69	0.51	0.18	0.29
1877	732	1.03	1.95	0.49	0.13	0.15
1887	806	1.02	1.87	0.50	0.12	0.12
1897	894	1.05	1.89	0.50	0.12	0.13
1899	953	1.03	1.76	0.51	0.12	0.13
1910	1119	1.06	1.59	0.53	0.11	0.11
1920	1230	1.07	1.60	0.54	0.12	0.13
1930	1544	1.13	1.92	0.56	0.16	0.21
1935	1724	1.16	1.83	0.57	0.17	0.23
1940	1869	1.17	1.91	0.58	0.18	0.28
1950	2209	1.23	2.02	0.60	0.23	0.40
1960	2350	1.31	2.29	0.63	0.28	0.57
1970	2712	1.43	3.81	0.65	0.34	0.84
1980	3197	1.40	5.00	0.64	0.36	0.98
1990	3522	1.38	5.31	0.65	0.37	1.02
2000	3809	1.32	4.99	0.64	0.36	0.99
2010	3726	1.33	4.97	0.64	0.36	1.00

Notes: These measures are derived from municipalities aggregated into 16 units based on municipal definitions in 1765. Total population includes the islands of Vieques and Culebra, but the measures of variation do not. The share of the five largest municipalities is calculated using the five most populated municipalities in a given census. The Gini coefficient is calculated using population density. The relative variance of log population density is the variance of the log of population density in year  $t$  divided by the variance of the log of population density in 2010.

uniformly distributed across municipalities and approaches one as population grows increasingly concentrated. The last measure, the relative variance of log population density, measures variance in proportional terms by dividing the variance of the natural log of population density for a census year by variance in 2010. Values greater than one indicate that population was more concentrated historically than currently.

Despite differences between the measures, they all generate a similar pattern of results. First, variation in the 18th century was much higher than modern variation, reflecting the fact that Puerto Rico was lightly populated in general, with a small settler colony heavily concentrated on the coast. Second, variation declined throughout the 19th century as population grew and filled in the interior of the island. Only in the 1827 census did variation fall below the modern level. The measures of variation continue to visibly decline until about 1877 and more or less plateau thereafter. In the late 19th and early 20th centuries, the population was the most evenly distributed across municipalities in recorded history. Third, variation then reversed direction, rising gradually through 1940 and rapidly from 1940 to 1970 or 1980, depending on the measure. Fourth, variation in municipal population density has been flat since about 1980, when the island had finished industrializing and the economy began an extended period of stagnation. These results are consistent with Vázquez Calzada (1978, 1984), who finds a U-shaped pattern of variation over time, with a trough in 1899, by estimating a Gini coefficient for disaggregated municipal data and seven regions, respectively. As occurred in the mainland US (Michaels et al., 2012) and elsewhere, the increase in variation coincided with the decline of agriculture (see Figure 4.2).

Whether 1897, 1899, or 1910 represents the lowest point in this series depends on the level of aggregation used, but  $F$ -tests fail to reject the null hypothesis that the variance of log population density between any pair of these years is equivalent. When using 1899 municipality definitions—the dataset which required the least amount of aggregation (see Tables C.1, C.2, and C.3 in Appendix C.2)— $F$ -tests suggest at the .10-level that variance was greater in 1930 than in 1899 and at the .05-level that variance was greater in 1940 than in 1899. These results are somewhat contingent on the level of aggregation but generally indicate a significant increase at least by 1940.  $F$ -tests

continue to detect rising variance between 1940 and 1970—specifically, variance is greater in 1960 than in 1940 and is greater in 1970 than in 1950, both significant at the .05-level—but there is no statistically significant difference between 1970 and later censuses. Therefore, these results indicate that municipal variation in population density began to increase shortly after the Spanish-American War, accelerated after World War II, and ceased around 1970.

#### **4.4.3 Correlates of Municipal Population Growth**

This subsection builds on the previous two by describing the patterns underlying changes in relative population density and variation in population density. Specifically, I use regressions to summarize the correlates of population density growth—or, equivalently, population growth, since the area of the municipality is constant—across censuses. Patterns of coefficients, including sign and magnitude, offer information about the relative growth of regions over time, providing context for changes discovered in the previous exercises.

The regressions are estimated as a repeated cross-section with population growth as the dependent variable. I use this model rather than a two-way fixed effects model (with year and municipality fixed effects and time-invariant correlates interacted with year fixed effects) because the purpose is to determine the correlates of population growth between censuses, irrespective of correlations in other years. If, for example, sugarcane cultivation were the best predictor of population growth throughout the sample but had the same relationship in each period, the repeated cross-section would reveal this important information whereas a fixed effects model would not. Each regression reveals the municipality characteristics that predict population growth in a given period, and tracking patterns of coefficients across regressions helps to see how the spatial distribution of economic activity changed over time.

I include only a few right-hand side variables to summarize broad patterns in population growth. First, to capture locational fundamentals, or natural endowments, I include measures of traditional crop production for the three crops that constituted the bulk of exports throughout history: sugar, coffee, and tobacco. Second, I include a measure of economic development to

capture the movement towards cities. The OLS model is as follows:

$$\Delta Population_{mt} = \beta_0 + \beta_1 Sugar_m + \beta_2 Tobacco_m + \beta_3 Development_m + \varepsilon_{mt} \quad (4.1)$$

where the dependent variable is annual population growth for municipality  $m$  from year  $t$  to the next census, multiplied by 100 for ease of exposition, i.e.  $\Delta Population_{mt} = 100 \times \frac{\ln Population_{m,t+1} - \ln Population_{mt}}{year_{t+1} - year_t}$ . Observations are weighted by population to summarize changes throughout the population in general, but unweighted results are reported in Appendix C.6. The analysis is based on municipality definitions from 1827, which includes 51 municipalities, and excludes the 1765, 1776, 1790, and 1800 censuses. Notably, the pattern of results is invariant to dataset used.

$Sugar_m$ ,  $Coffee_m$ , and  $Tobacco_m$  are dummies equal to one if a municipality is in that crop region, as defined by Ayala and Bernabe (2007), and zero otherwise.  $Coffee_m$  is the omitted category. All municipalities fall into one of the three crop regions, as all municipalities historically included farmland. Results for other crop variables, described in Appendix C.1, are similar and are therefore relegated to Appendix C.6. The only major difference in the results for crop variables is that population growth after 1899 tends to be *negatively* correlated with the share of land dedicated to sugar cultivation in 1910. However, this result is not found in unweighted regressions (not shown) or when evaluating net migration in Appendix C.3. In any case, crop production measures are more easily interpretable than geographic characteristics, and crop dummies are preferred to the 1910 crop shares because they do not rely on an outcome from one census.

In my preferred specification,  $Development_m$  is equal to the share of the labor force employed in agriculture in 1899, standardized to have a mean of zero and a standard deviation of one. The agricultural employment share is an important measure of development because urbanization and economic modernization are closely related to structural transformation away from agriculture (Davis and Weinstein, 2002; Henderson, 2010; Michaels et al., 2012). I assume that relatively more developed municipalities in 1899 had previously been more developed and continued to be so, at least in the short-term. This assumption is most reasonable in the years proximate to 1899, so I

focus on results for those years. The assumption is also plausible for much of the 19th century since the municipalities with the lowest share of their labor forces in agriculture were by and large either the seats of Spanish administration or nearby.

While a time-varying development variable may be more desirable, such data are not available prior to the first US census in 1899. Besides, for robustness I limit the sample to years in which municipal-level literacy data are available, from 1887 to 2000, and estimate the regression using *contemporaneous* literacy rates, and the pattern of results is unchanged (see Figure 4.7). The pattern of results is also similar when the 1899 development variable is replaced by a 2013 socioeconomic development index or a later value of the development variable, such as agricultural employment in 1935 or 1960. There is persistence in the relative level of development by municipality at least from early 20th century onwards, so there is not much to gain in the post-1899 analysis by using time-varying development data.

The results of Equation 4.1 are presented in Table 4.3. Population growth was negatively associated with the sugar region, relative to the coffee region, throughout the 19th century, particularly after 1860. The comparatively sluggish growth of sugar municipalities coincides with the stagnation of the sugar industry in the 1840s, largely the result of a failure to modernize mills at a time of increasing international competition including the ascent of subsidized European beet sugar. Coffee production expanded in response to rising international prices in the late 1880s and early 1890s, and these results suggest that the expansion was helped by a movement of labor from the sugar- to coffee-producing municipalities: from 1887 to 1899, specializing in sugar rather than coffee is associated with 0.8 percentage points lower annual population growth, significant at the .05-level. There is no consistent pattern of coefficient estimates for *Tobacco<sub>m</sub>* during the 19th century.

Table 4.3: Correlates of Population Growth

	1827	1846	1860	1877	1887	1899	1910	1920
Agricultural Employment, 1899	0.21** (0.10)	0.01 (0.18)	-0.03 (0.09)	-0.05 (0.08)	-0.05 (0.10)	-0.43*** (0.09)	-0.28* (0.15)	-0.63*** (0.11)
Sugar	-0.22 (0.43)	-0.56 (0.48)	-0.59* (0.31)	-0.36 (0.31)	-0.77** (0.30)	1.27*** (0.27)	0.47* (0.28)	0.56** (0.26)
Tobacco	0.42 (0.74)	-0.27 (0.87)	0.10 (0.51)	0.30 (0.69)	-0.60 (0.40)	1.66*** (0.57)	2.03*** (0.41)	1.87*** (0.56)
Constant	2.49*** (0.37)	2.13*** (0.42)	1.59*** (0.32)	1.10*** (0.26)	1.82*** (0.21)	0.36 (0.25)	0.92*** (0.21)	0.88*** (0.19)
R-squared	0.09	0.02	0.08	0.05	0.11	0.54	0.41	0.62

	1930	1940	1950	1960	1970	1980	1990	2000
Agricultural Employment, 1899	-0.44*** (0.11)	-0.67*** (0.12)	-0.52*** (0.09)	-0.12 (0.10)	0.42*** (0.14)	0.20*** (0.04)	0.32*** (0.03)	0.23*** (0.03)
Sugar	0.30 (0.28)	0.78** (0.30)	0.60* (0.32)	1.66*** (0.46)	0.77 (0.59)	0.61*** (0.15)	0.37*** (0.15)	0.49*** (0.13)
Tobacco	-0.26 (0.38)	0.68* (0.35)	1.12** (0.46)	1.03 (0.68)	-0.69 (0.54)	0.72*** (0.25)	-0.09 (0.30)	0.57*** (0.17)
Constant	1.45*** (0.21)	0.56** (0.24)	-0.40* (0.23)	-0.04 (0.25)	1.53*** (0.46)	0.66*** (0.11)	0.79*** (0.11)	-0.44*** (0.10)
R-squared	0.65	0.72	0.53	0.20	0.30	0.41	0.49	0.39

Notes: The dependent variable is the annual population growth rate between censuses, expressed as a natural log. Each column represents a separate regression based on Equation 2.1, each with 51 observations. Variables are described in the data appendix. Observations are weighted by municipal population. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

The early 20th century is marked by drastic changes in the correlations between population growth and crop regions, as population growth was strongly positively correlated with sugar and tobacco cultivation. These changes provides context for the radical change in relative population density after 1899 as described in Section 4.4.1. Between 1899 and 1910, the sugar region is associated with 1.3 percentage points more annual population growth relative to the coffee region, and the tobacco region is associated with 1.7 percentage points more growth. Sugar cultivation continues to be positively correlated with population growth throughout the rest of the sample, and the coefficient estimate is typically statistically significant but smaller in magnitude. Notably, the correlation is not significant in the 1930s, despite the fact that the sugar industry fared well relative to the coffee industry during the Great Depression. Tobacco cultivation remains positively correlated with population growth until 1930, around the demise of the tobacco industry during the Great Depression. Subsequently, the region generally experienced higher growth again, but this relationship is not robust to using other measures of tobacco production (see Table C.6). In any case, postwar correlations do not reflect the migration of agricultural laborers given the decline of agriculture.

In summary, population growth was positively correlated with crop production—coffee in the late 19th century, and sugarcane and tobacco in the early 20th century—in the preindustrial era and thereafter continued to favor the areas of more rapid growth prior to industrialization. This pattern is consistent with the hybrid theory of economic geography proposed by Davis and Weinstein (2002), where industrialization reinforces the spatial equilibrium put in place by locational fundamentals in the preindustrial era. This is noteworthy given that regional policies were designed to attract economic activity to less developed areas (Guimaraes et al., 1998).

Turning to the development variable, population growth is not significantly correlated with agricultural employment in the 19th century other than the period from 1827 to 1846, during which there is a positive correlation. This suddenly changes between 1899 and 1910, when a one standard deviation increase in agricultural employment corresponded to a decrease of 0.3 percentage points in annual population growth. The negative correlation between population growth and agricultural

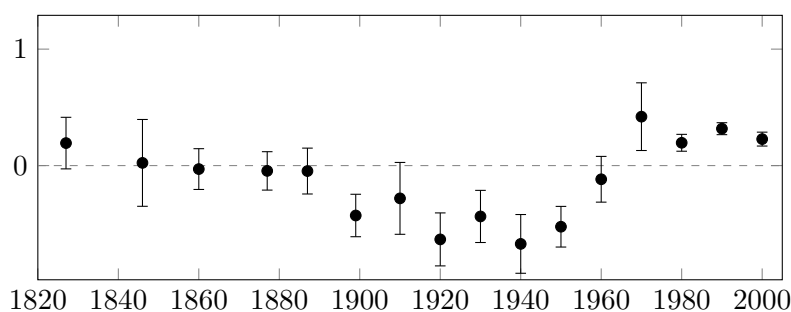
employment remains until 1960 and is usually statistically significant at the .01-level. From 1970 onwards, during a prolonged economic slowdown, the coefficient on agricultural employment is positive and statistically significant. The positive relationship is explained mostly by the fact that three of the island's largest cities—San Juan, Ponce, and Mayagüez—experienced slow and even negative growth in recent decades. Looking at the modern socioeconomic development index—which is the most appropriate measure of development for the postwar era—instead of historical agricultural employment, population growth has been unrelated to development since 1970 (see Figure 4.6). Over this period, the island has undergone suburbanization, with particularly high growth in the municipalities outside of San Juan.

The lack of statistical significance of agricultural employment from 1960 to 1970 is at first surprising because explosive economic growth did not cease until the 1970s. However, this result is driven by the rapid growth of several municipalities which border San Juan and did not have low agricultural employment shares in 1899. Clearly, however, the growth of these municipalities should be interpreted as an outgrowth of the island's most developed city. A one standard deviation increase in *contemporaneous* agricultural employment is associated with 0.5 percentage point lower annual population growth from 1960 to 1970 ( $p = 0.00$ ) and essentially uncorrelated with population growth from 1970 to 1980 ( $p = 0.75$ ) (not shown). Population growth is also significantly positively associated with contemporaneous literacy rates from 1899 to 1970 (see Figure 4.7), and there is a strong positive correlation between growth from 1960 to 1970 and the modern socioeconomic index (see Figure 4.6).

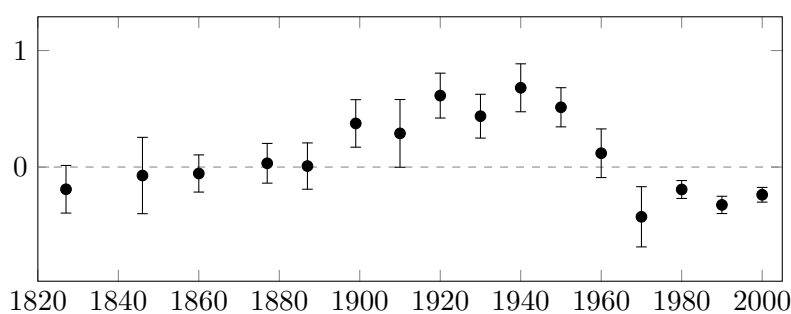
As shown in Figure 4.6, the main results generally hold regardless of the development variable used: from 1899 to 1970, population growth was positively tied to literacy rates and manufacturing employment. Strikingly, the pattern of coefficients is quite similar for the 2013 Socioeconomic Development Index, illustrating that there is a high degree of persistence in relative development of municipalities over the past century and supporting the use of the 1899 variables. For all development variables the pattern of coefficients is similar albeit less pronounced when observations are unweighted, so the results generalize to municipalities as well as the island population (see Figure

Figure 4.6: Development Variables and Population Growth by Census Year

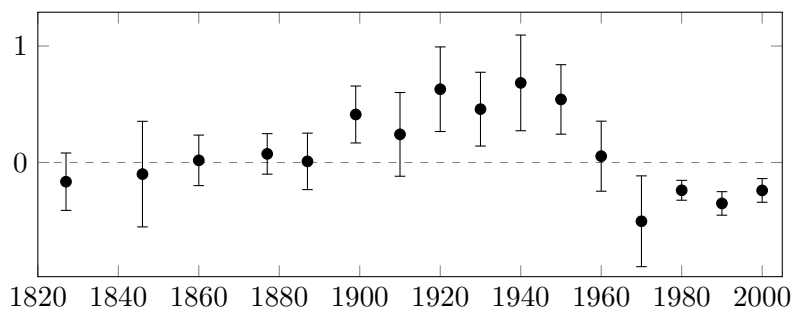
(a) 1899 Agricultural Employment



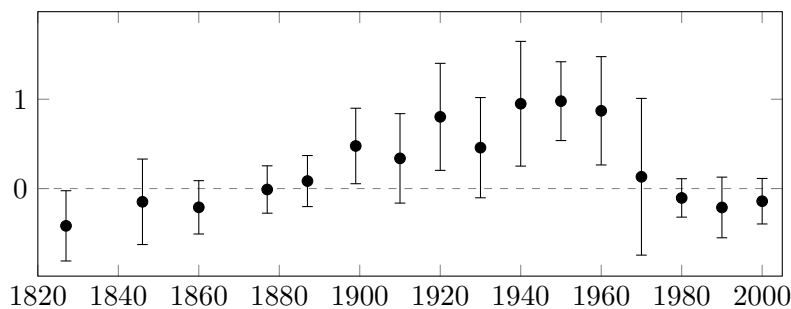
(b) 1899 Literacy Rate



(c) 1899 Manufacturing Employment



(d) 2013 Socioeconomic Development Index

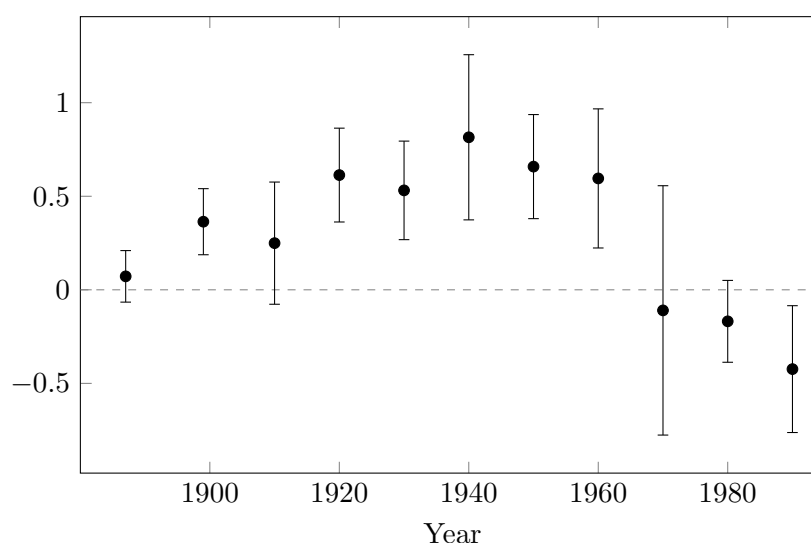


Notes: Regression coefficients and 95 percent confidence intervals come from Equation 2.1, with the population growth rate as the dependent variable. Coefficients correspond to the right-hand side variable shown at the top of each panel. Observations are weighted by population. Regressions include 51 observations, based on municipalities aggregated for analysis starting in 1827.

C.1 in Appendix C.5).

Hence, more developed municipalities grew at a faster pace for the first seven decades of the 20th century. This trend ceased after 1970, when the economy stagnated and convergence to US income halted. My results complement Ayala and Bergad (2020), who emphasize that population growth was greater in urban areas than rural areas in each region and in each decade between 1899 and 1940 and argue that population growth in urban areas was clearly tied to the economic expansion taking place in each region. The movement out of areas heavily dependent on agriculture coinciding with increasing spatial concentration of population is reminiscent of the interrelated processes of urbanization and structural transformation in the US (Michaels et al., 2012).

Figure 4.7: Population Growth and Contemporaneous Municipal Literacy Rates



Notes: The regression coefficients and 95 percent confidence intervals come from regressions with municipal population growth rate as the dependent variable, estimated as a repeated cross-section. Coefficients correspond to literacy in year  $t$ , standardized to have a mean of zero and a standard deviation of one. Crop dummies are included in each regression as controls. Observations are weighted by municipal population. Each regression includes 68 observations, based on municipalities aggregated for analysis starting in 1887.

Finally, it is plausible that observed differences in city growth may be driven by differences in natality and mortality. To account for births and deaths, in Appendix C.3 I evaluate net migration by municipality from Vázquez Calzada (1978) for 1899 to 1930 and each decade from the 1930s

through the 1960s. The patterns that emerge are consistent with those previously discussed: in the early 20th century, net migration was positively correlated with sugar and tobacco and negatively correlated with agricultural employment and coffee.

#### 4.5 Urbanization in Comparative Perspective

Of the periods of change described in Section 4.4, none stands out more clearly than the early 20th century. After 1899, the spatial distribution of population quickly approached the modern distribution. Variation in municipal population density began to increase early in the century, driven by the growth of Greater San Juan. From 1899 to 1970, labor left agricultural areas for more developed municipalities. Taken together, these results suggest that urbanization began around the start of the 20th century.

That urbanization commenced around 1900 is an important finding because past research tends to treat post-WWII industrialization as a sharp break from a stagnant past (e.g., Baer, 1959). Taylor (1952, p. 42), for example, writes that Puerto Rico “languished as a typical underdeveloped area” before the postwar boom. Yet, the spatial trends exhibited during the postwar “miracle” were a mere continuation of trends originating at the turn of the century. Equating urbanization with economic growth cannot be supported with my data and is unnecessary given available GDP data. Still, urbanization is a key feature of the development process. My results indicate that Puerto Rico’s impressive socioeconomic transformation in the 20th century began, contrary to most accounts, before World War II.

Was urbanization in Puerto Rico merely part of a regional trend? Placing Puerto Rico in a regional context helps to understand the forces driving its urbanization and contributes to the economic history of the Latin America and the Caribbean more broadly. Below I more formally measure urbanization in Puerto Rico and make comparisons to nearby countries.

Urbanization is difficult to define and measure. Political boundaries often do not reflect economic realities. Measures of urbanization differ over time and across countries. In Puerto Rico, the official urbanization rate, defined as the share of the population living in places of 2,500 or more

inhabitants, grew from 14.6 percent in 1899 to 30.3 in 1940 and 40.5 in 1950 (Vázquez Calzada, 1978). But it is difficult to know to what extent the change in urbanization is due to general population growth, and these figures cannot be measured before 1899 because Spanish data are not available below the level of the municipality.

I construct a time series of an urbanization index using definitions of Metropolitan Areas (MA) in 1950 delimited according to common criteria by Kingsley Davis and International Urban Research (1959). MAs are meant to include the territory surrounding the city if it represents a continuous extension of the city or if the people living there are predominately connected with the city in an intimate day-to-day economic and occupational sense:

[The MA is an] area with 100,000 or more inhabitants, containing at least one city (continuous urban area) with 50,000 or more inhabitants and those administrative divisions contiguous to the city (or to the continuous urban area) which meet certain requirements as to metropolitan character. (p. 25)

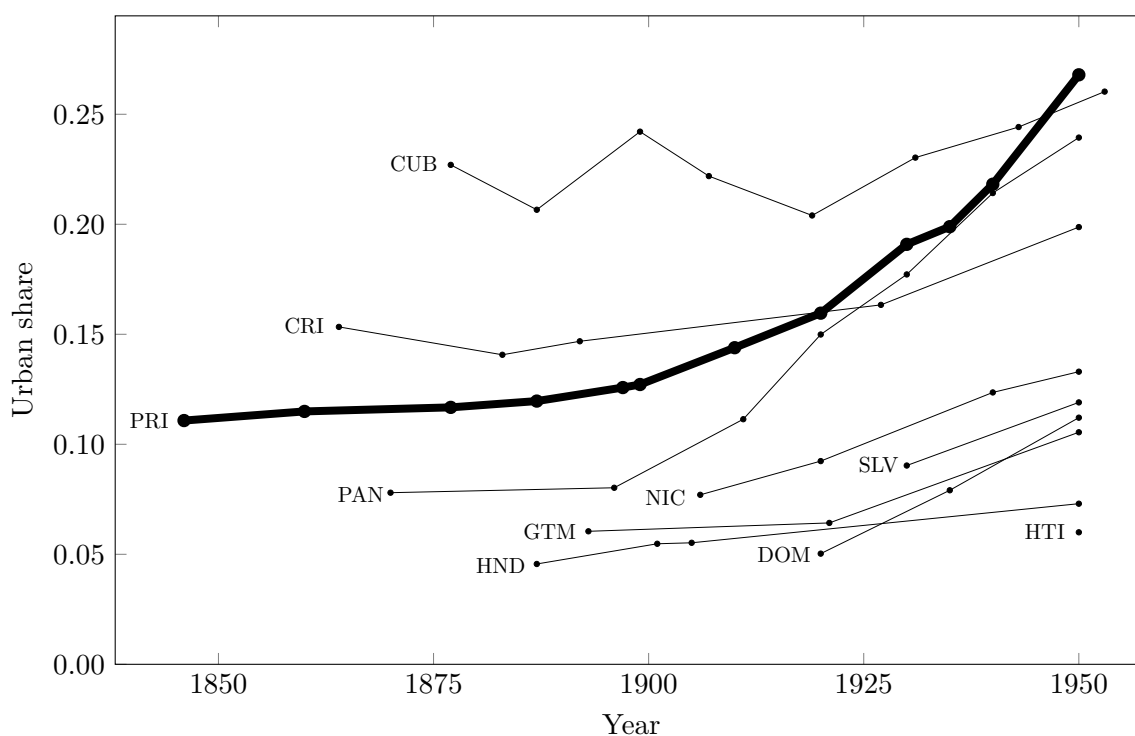
The time series represents the share of the population in a census year residing in MAs according to the 1950 definition. Therefore, changes in the MA over time do not reflect the rise of new MAs but, rather, the growth of areas which were and would become the main centers of population by 1950. Puerto Rico has two MAs, Ponce (municipality of Ponce) and San Juan (municipalities of San Juan, Bayamón, Cataño, Guaynabo, and Río Piedras).

This index allows for comparison across countries and over time because MAs are approximately standardized and consist of only a few administrative units. Carrying out the exercises in Section 4.4 for a sample of countries is impractical because differences in administrative units both across countries and within countries over time are substantial and changes in administrative units over time require very different degrees of aggregation between countries to harmonize boundaries. Moreover, nearby countries generally did not carry out censuses consistently, leaving large gaps between data points, and did not publish much information about changes to administrative divisions between censuses.

In Figure 4.8, I compare urbanization in Puerto Rico to countries in the Caribbean and Central America because of cultural, political, geographic, demographic, and economic similarities.

The component units of each metropolitan area by country are listed in Appendix C.4. These countries are also of comparable size, including land area and population, and most have just one or two MAs, like Puerto Rico. Since the index represents the share of the population residing in the administrative divisions that would constitute MAs in 1950, cross-country comparison of levels is appropriate for 1950 but must be done with caution in years prior, as the geographical area of the MA is a function of urbanization in 1950. Cross-country comparisons for years prior to 1950 should focus more on patterns of growth.

Figure 4.8: An Urbanization Index for Central America and the Caribbean



Notes: The urbanization index represents the share of the population residing in the administrative divisions that constituted part of a Metropolitan Area in 1950, as defined by International Urban Research (1959). See Section 4.5 for more details.

Sources: Author's calculations based on national population censuses.

Between 1846 and 1899, the share of the population in MAs in Puerto Rico increased slightly, from 11.1 to 12.7 percent. The share then rose to 14.4 percent in 1910, a larger increase in one decade than had occurred in more than a half century prior, and continued upwards to 21.8 percent

in 1940. By 1950, the urbanization rate in Puerto Rico reached 26.8 percent, the highest rate in Latin America outside of the Southern Cone (43.8 in Argentina, 32.7 in Uruguay, and 29.6 in Chile), just ahead of Venezuela (26.2) and Cuba (26.0) (Browning, 1958).

The timing of urbanization in Puerto Rico is unique. Latin America and the Caribbean have a long history of urban dominance from the beginning of colonization, but urban growth was slow and halting for centuries until the end of World War II (Boswell, 1978; Browning, 1958). Haiti, the Dominican Republic, and the small republics of Central America experienced little urbanization before at least 1930 (Hardoy, 1975). Panama is an exception, as Panama City rapidly grew during the construction of the canal starting in 1904. Cuba, too, exhibits an early upward trend, driven by the prosperity of the sugar industry, but was already highly urbanized in 1900 (Dyer, 1957). Across the region, however, there is no general upward trend in Figure 4.8 until the end of the sample. Urbanization in Puerto Rico continued after World War II but commenced much earlier. Notably, the population of San Juan (including Río Piedras, annexed in 1951) grew from 43,276 in 1899 to 237,537 in 1940, outpacing the growth rate of the island's population by a factor of 4.7. Vázquez Calzada (1978) estimates a net rural-urban migration of 326,000 from 1899 to 1940 and 703,000 from 1940 to 1970.

The fact that urbanization in Puerto Rico began shortly after US annexation motivates further research. Urbanization in most of Central America and the Caribbean began decades later, and the two countries in the region that also experienced significant urban growth in the early 20th century were Cuba and Panama, the countries most intimately tied to the United States. Vázquez Calzada (1984) claims that the changes in the geographic distribution of the population, including the rapid growth of San Juan and the northeast coast, were caused principally by the economic changes resulting from US annexation. More recently, Ayala and Bergad (2020) attribute early urbanization to the export-driven economy, which created a domestic market for a variety of goods and services, as well as the expansion of government bureaucracies. The authors also emphasize the employment opportunities for women in cities in the new developing market economy that emerged after 1898.

The prevailing view in the literature holds that US annexation led to stagnant or even worsening economic conditions. Yet, recent work on the period from 1900 to 1940 shows large improvements—among the largest in the region—in GDP per capita (de Jesús Toro, 1982; Devereux, 2019), exports (Bulmer-Thomas, 2003), life expectancy (see Chapter 2), literacy and primary enrollment (Nuñez, 2005), adult height (see Chapter 5), and transport infrastructure (see Chapter 3). Likewise, there are studies finding a positive effect of US influence in the early 20th century on development in Hawaii after US annexation (Burdekin and Laney, 2008) and the Circum-Caribbean in response to the Roosevelt Corollary (Mitchener and Weidenmier, 2005). US involvement also ended presidential turnover and supported foreign investment in Haiti (Palsson, 2020). This paper is consistent with these studies and warrants a reconsideration of the effects of US annexation on Puerto Rican development.

There are several competing explanations for Puerto Rico's early urbanization. First, there is a pervasive myth that US sugar corporations dispossessed Puerto Rican farmers of land, but several studies show that rural landownership expanded after 1899 (Ayala and Bergad, 2002, 2020; Solá, 2011). Second, many studies blame US policy for devastating the coffee region, but coffee exports recovered to pre-annexation levels after the destruction of Hurricane San Ciriaco in 1899, and US authorities tried to help farmers resurrect the region (Ayala and Bergad, 2020). Third, population might have gravitated towards places where the United States established social programs and expanded government bureaucracies. However, just 1.2 and 1.0 percent of the labor force was employed in public administration in the 1910 and 1920 censuses, respectively, and just 7.1 and 4.1 percent in San Juan (Ruggles et al., 2018). Puerto Rico had a small government before the New Deal, and policy was explicitly designed to provide rural areas with schools (Bary, 1923), roads (see Chapter 3), and health services (see Chapter 2).

## 4.6 Conclusion

The purpose of this paper is to comment on economic development in Puerto Rico in the long-run by analyzing trends in the spatial distribution of population. Spatial population trends

provide information about broad patterns of economic activity over time. I correct for changes to municipal boundaries by grouping municipalities into larger units on the basis of historical configurations, allowing for quantitative comparisons across censuses.

Early censuses show a small population clustered on the coast. The population then rapidly increased with immigration from Europe and Latin America and spread throughout the interior of the island, especially the coffee region in the western highlands, in the 19th century. The early 20th century was marked by higher population growth in areas of sugar and tobacco cultivation on the coast and eastern interior, as well as gradual urbanization. Municipal population densities began to diverge, increasingly so between 1940 and 1970. After World War II, the population concentrated in Greater San Juan, and population growth was highest in the municipalities outside of San Juan proper near the end of the century during a period of suburbanization. These patterns are consistent with past descriptive work (Bothwell, 1937; Picó, 1974; Vázquez Calzada, 1978, 1984) and short-run quantitative studies (Boswell, 1978; Jackson, 1980; Stone, 1965; Vaughan and Schwirian, 1979). The conclusion that urbanization began around 1900 is consistent with a discussion of economic transformation and demographic change in Ayala and Bergad (2020), and the fact that most of Central America and the Caribbean lagged by several decades motivates further research regarding the causes of Puerto Rico's early urbanization.

## Chapter 5

### Economic Development in Puerto Rico After US Annexation: Anthropometric Evidence

A version of this chapter was published in *Economics & Human Biology* in August 2020.

#### 5.1 Introduction

Puerto Rico is the most populous unincorporated territory of the United States, and Puerto Ricans have been US citizens for more than a century. After World War II, Puerto Rico experienced one of the world's fastest growth rates for GDP per capita and labor productivity (Baumol and Wolff, 1996), became in 1967 the first country in Latin America to be classified as a “developed nation” by the World Bank (Duchesne, 2000), and remains one of the richest economies in Latin America. Nonetheless, the economic value of political association with the US has been historically, and is presently, a matter of considerable debate. US annexation in 1898, in particular, is perhaps the most controversial topic in Puerto Rican economic history. Scholars of early 20th century Puerto Rico have reached a broad range of conclusions about US annexation, from “extreme denunciations of US imperialism and its purported ruthless exploitation and impoverishment of the island...” to “unabashed praises of the US invasion for having saved Puerto Rico from the tyranny of Spanish colonialism, for building a modern economic infrastructure, and for eventually ushering in an era of democracy, relative prosperity, and social mobility...” (Ayala and Bergad, 2002, p. 65).

A key challenge to resolving the academic debate on conditions in early 20th century Puerto Rico has been a lack of quantitative evidence to measure living standards. In this paper, I draw

on anthropometric evidence from three surveys of Puerto Rican men to comment on economic development. I exploit cross-sectional variation in birth year to identify differences in average height over time and compare age cohorts across surveys where possible. Increases in height have been found to be associated with rising living standards, and height data have been used in various other historical contexts to comment on changes in material living conditions under colonialism (Brennan et al., 1997; Komlos, 2001; Moradi, 2009; Morgan and Liu, 2001; Olds, 2001). Likewise, I show that there was a significant rise in the height of Puerto Rican men following US annexation: adult men born in 1940 measure about 4.2 cm. taller than those born a decade prior to US annexation. Additionally, I show that Puerto Ricans at mid-century were among the tallest Latin Americans outside of Argentina and Uruguay.

Lacking an adequate control group and sufficient observations predating US rule, I do not argue that US annexation is responsible for changes across birth cohorts; after all, the 20th century witnessed improvements in living standards and stature across the West (NCD Risk Factor Collaboration (NCD-RisC), 2016).<sup>1</sup> Rather, I more narrowly contend that widespread beliefs about impoverishment caused by US policy, or even stagnant economic conditions in the immediate aftermath of annexation, are incorrect. Such claims have been shaped in large part by failing to consider the realities of Spanish colonialism—and, therefore, blaming all problems on the US—and by observing that living standards in Puerto Rico were much lower than those of the US and Western Europe without properly comparing Puerto Rico’s present conditions with the conditions that prevailed at the commencement of US rule (Ayala and Bergad, 2002; Chamber of Commerce of Puerto Rico, 1945). My findings are supportive of a marked improvement in the biological standard of living and are not attributable to selection into the sample or shrinking in old age. Of course, this does not imply that US rule was good, nor does it excuse injustices committed against Puerto

---

<sup>1</sup> To my knowledge, the only other available source of anthropometric data for Puerto Ricans born under late Spanish rule is Fleagle (1917), a survey of students at the University of Puerto Rico. Comparison to that study is of little value in the present analysis due to the unavailability of microdata, the small size of the sample, the fact that birth years overlap with those in the PRHHP sample and thus do not provide much additional information, and, most importantly, that there must be an extremely high degree of selection into the sample. Note that in the PRHHP, only 8 percent of participants reported having at least some college education, and only 4 percent of those born in 1900 or earlier.

Ricans by the US government; rather, this paper focuses on material living standards in response to previous research.

Evidence of an improvement in the material well-being of Puerto Ricans stands in stark contrast to the prevailing view that US rule led to stagnant or even worsening conditions for most Puerto Ricans before World War II. For instance, D'Estefano (1968, p. 21) alleges that "Yankee colonialism" had produced "unemployment, exploitation, disease, hunger, and 400,000 illiterates." Renowned historian Arturo Morales Carrión (1983, p. 153) describes Puerto Rico in this era as a "poorhouse... foundering in despair", while political scientist Gordon Lewis (1963, pp. 89-90) insists that the revamped economy worked for the "interests of the colonial producer and not for those of the consumer", most of whom were reduced to "a starvation diet of the very simplest of ground provisions." Even the most widely cited economic history of Puerto Rico contends that there is "little evidence" of "improvements in the standard of living of the mass of working and nonworking people" (Dietz, 1986, p. 133).

My results also contradict the conclusions of Godoy et al. (2007), the only empirically grounded study with a somewhat negative appraisal of changes in living standards in the early 20th century. Despite using data from the same 1965 health survey analyzed in this paper, the authors find no statistically significant increase in height, which they ascribe to "the offsetting role of improved health and a stagnant rural economy during the first half of the 20th century" (p. 82). I argue that their result is the product of an overspecified regression model, which includes birth year fixed effects along with an age variable, generating severe collinearity and without adding meaningful information to the model. Furthermore, they include educational attainment and an urban/rural dummy as explanatory variables, diminishing the magnitude of the estimated effect of birth year. Including these variables purges the birth cohort fixed effects of variation attributable to rising levels of education and urbanization, which are correlated with time and are part of, not distinct from, the development process. A simple correction renders evidence of a clear rise in height and, therefore, the biological standard of living.

The results in this paper are, however, consistent with recent quantitative analyses that

suggest that the early 20th century was a period of economic growth in Puerto Rico. Most importantly, Devereux (2019) and de Jesús Toro (1982) estimate that real income in 1940 was about two and a half times higher than in 1900. The results hold after adjusting GDP estimates for fluctuations in the terms of trade and are corroborated by per capita consumption data: food intake increased by more than 50 percent, and clothing consumption doubled. Additional evidence in support of economic growth is presented in Chapter 4, which finds that economic activity began concentrating in urban areas at the turn of the 20th century, and municipal growth was positively correlated with several indicators of development from 1899 to 1970. Height provides additional information about development because it captures how living conditions changed throughout the entire population; economic growth can fail to translate to improvements in living conditions for vast sections of the population if income inequality rises as well.

In what follows, I first provide a broad overview of economic changes in the early 20th century as well as interventions made under US rule that may have affected the standard of living. I subsequently evaluate height data for Puerto Rican men from three surveys and determine that the economic standard of living rose substantially in the immediate aftermath of US annexation. I also show that Puerto Ricans at mid-century were among the tallest Latin Americans outside of Argentina and Uruguay and were nearly as tall as white Cubans, at which point Cuba was a prosperous, middle-income country. Finally, I summarize and conclude.

## 5.2 Puerto Rico in the Early 20th Century

Prior to US annexation, Puerto Rico had been a Spanish colony for nearly four centuries. Madrid valued the island as a naval outpost, leading to an overarching policy of neglect (Dietz, 1986). Despite liberalizing reforms in the 19th century, Spain continued to impede Puerto Rican economic development by severely restricting credit and the importation of industrial machinery. External trade volume grew roughly in proportion to population in the 19th century, suggesting little economic progress (US Tariff Commission, 1943).

The US annexed Puerto Rico at the conclusion of the Spanish-American War in 1898. The

population was predominately rural and illiterate, and most adults lacked even necessities (Davis, 1900). Malnutrition was quite severe, leaving many susceptible to intestinal disorders (Garver and Fincher, 1945). Infectious diseases, such as smallpox, were rampant (Berbusse, 1966), and hookworm infection was “so common as almost to justify the claim that it was universal” (Clark, 1930, p. xviii). Streams often served both as a source of drinking water as well as common sewers (Governor of Puerto Rico, 1901).

The US responded with historic investments in public health and education during the period of direct rule from 1898 to 1952. Just weeks after the Treaty of Paris, the US military instituted a compulsory smallpox vaccination campaign, inoculating 790,000 people in three months (Trujillo-Pagan, 2013). US officials also helped to update water provision, construct sewer systems, launch health education campaigns, and build hospitals (Espinosa, 2000). Before WWII, Puerto Rico had by far the highest health and sanitation expenditures in relation to land area in the Circum-Caribbean and spent roughly equal amounts in per capita terms as Cuba and Venezuela (Smith, 1937). These changes should have contributed to increased stature since Silventoinen (2003) concludes that exposure to disease is one of the most important non-genetic factors determining height, while Hatton (2014) finds that the most important proximate source of increasing height in Europe from the mid-nineteenth century to the 1980s was the improving disease environment as reflected by the fall in infant mortality. In Puerto Rico, the available evidence indicates that infant mortality was extraordinarily high and fairly stable throughout the late 19th century and dropped markedly in the first 15 years of the 20th century. The infant mortality rate per 1,000 births was about 244 in the decade preceding the Spanish-American War but fell to 152 for the period 1910-1914 (Vázquez Calzada, 1978). The decline in infant mortality then stalled before recommencing in the late 1920s and dropping to 68 by mid-century, the lowest rate in Latin America outside of Uruguay (Molina and Noam, 1964).

Free public education drastically expanded such that the percentage of children ages 7 to 13 who attended school rose from 11.6 percent in 1899 to 45.3 percent in 1910, and all the way to 66.8 by 1940. Only 18.3 percent of adults could read and write in 1899—the lowest rate in Latin

America other than Bolivia and Guatemala (Nuñez, 2005)—but this figure jumped to 74.4 percent by 1950. Puerto Rico soon had the highest rate of school attendance in the Circum-Caribbean (Smith, 1937). Since height has been found to be positively associated with education, the spread of public education also suggests an increase in stature (Silventoinen, 2003).<sup>2</sup> Regardless of changes in income, living standards broadly interpreted certainly improved (Prados de la Escosura, 2015).

Large public investments were also made in infrastructure, such that by the end of WWI there were “thousands of kilometers of roads, irrigation systems, dams and hydroelectric projects, railway and tramways, telegraph and telephone systems, and ports. . .” (Cabán, 2002, p. 132). Before WWII, Puerto Rico had by far the highest length of highways and railways, respectively, by area in the Circum-Caribbean (Smith, 1937), and investments disproportionately benefited those areas which formerly lacked modern facilities (see Chapter 3).

Most importantly, Devereux (2019) and de Jesús Toro (1982) estimate annual per capita income growth of around 2 or 3 percent in the first four decades of US rule. Coffee production gave way to more capital-intensive sugar cane and tobacco cultivation and manufacturing, fueled by unprecedented quantities of foreign investment following changes to the commercial code and the imposition of US tariffs. From 1899 to 1940, employment in agriculture as a percentage of all jobs fell from 62.8 to 45.5 percent, and employment in industry grew from 10.3 to 27.3 percent. Consistent with this trend, urbanization began sharply after US annexation (see Chapter 4).

To summarize, Puerto Ricans in the early 20th century enjoyed rising incomes and greater access to health and educational services, and these changes should have led to increases in stature. Yet, the prevailing view historically and presently has been that most of the benefits of increased output and external trade were captured by foreign shareholders, as profits were “either invested or consumed in the United States” and companies “extract[ed] more value out of workers by lowering wages”, generating a “profound deterioration in the conditions of Puerto Rican society” (Cabán, 1999, p. 251). The alleged displacement of small independent growers by American absentee sugar

---

<sup>2</sup> As mentioned by Godoy et al. (2007), education may affect height by increasing adult income and by teaching children better health habits. For instance, teachers in Puerto Rico instructed pupils how to avoid contracting uncinariasis (Clark, 1930, p. 66).

corporations reduced Puerto Ricans to “virtual economic serfdom” and forced them to import and pay a higher price for all their food staples, leading to “more widespread misery and destitution... in Puerto Rico than at any previous time in its history” as quoted by US Secretary of the Interior Harold Ickes in 1935 (Dietz, 1986, p. 132).

Besides the fact that land tenure did not become more concentrated in fewer hands under early US rule (Ayala and Bergad, 2002), consumption data overwhelmingly suggest that material living conditions improved for the mass of Puerto Ricans. In spite of increases in food prices, food intake increased by more than 50 percent from the mid-1900s to 1940 (Devereux, 2019). Puerto Ricans enjoyed many new products and consumed far more staple foods in 1940 than they had in 1900. For instance, per capita consumption of milk was more than 12 times what it was in 1900 by the most conservative estimates (Chamber of Commerce of Puerto Rico, 1945). Though poor by American or Western European standards, Puerto Rico was much richer by WWII than it had been at the time of US annexation.

### 5.3 Data

I analyze data on the height of Puerto Rican men from three surveys to comment on economic development in the early 20th century. The underlying logic for using these data has been well-established in the economics literature: height is positively correlated with the economic standard of living, broadly defined, and may serve as a proxy for the “biological standard of living” when more conventional indicators are unavailable, particularly in historical analysis (Komlos and Baten, 1998; Steckel, 1995, 2009). Height is heavily influenced by nutrition and disease during childhood, with the two most important periods of growth being from conception to two years and adolescence before the onset of puberty (Perkins et al., 2016). If conditions improved over time, later birth cohorts should be taller on average.<sup>3</sup>

---

<sup>3</sup> Although height and income often move in parallel, it is also possible for height to decline while income rises, as in the “Antebellum puzzle.” However, as discussed by Steckel (2009), countries that industrialized after the revolution in public health inspired by the germ theory of disease either had minor height cycles or experienced sustained upward trends. The post-annexation economic expansion in Puerto Rico coincided with large investments in public health.

### 5.3.1 Puerto Rico Heart Health Program (PRHHP)

The primary source of data in this paper is the Puerto Rico Heart Health Program (PRHHP), conducted in Puerto Rico between 1965 and 1980 by the National Heart, Lung, and Blood Institute to investigate morbidity and mortality from cardiovascular heart disease in men (García-Palmieri et al., 2002).<sup>4</sup> The PRHHP is the largest of the three samples and contains, by far, the most measurements of individuals born before or immediately after US annexation. The study included a total of 9,824 men, 6,848 of whom came from three municipalities (Bayamón, Guaynabo, and Carolina) of the San Juan Metropolitan Area, the island's largest urban area, and 2,976 of whom came from four rural, mountainous communities (Barranquitas, Cidra, Comerío, and Naranjito) just outside of San Juan. Communities were selected which, according to the 1960 census, contained approximately the same ratio of urban/rural population as the general population. A house-to-house census of the enumeration districts was made by the same personnel who participated in the decennial US census. All men of age in the enumerated area were invited to participate, and roughly 80 percent accepted in both the urban and rural areas. Three follow-up interviews were carried out, ending in 1980, but height was only measured at the first interview. Height, rounded to the nearest inch, is reported for 99.8 percent of participants. Men shorter than 58 inches were assigned a height of 58 inches, and those taller than 72 inches were assigned a height of 72.

The PRHHP was intended to study men ages 45 to 64 but ultimately included some individuals 35-44 and 65-79 years of age as well. Hence, the survey data include men born between 1886 and 1930, allowing me to evaluate changes in the standard of living in the late 19th and early 20th centuries. The public-use data for the PRHHP hide the birth year and age of the participant and instead bucket individuals into 5-year age cohorts. The distribution of participants by age cohort is displayed in Table 5.1. Clearly the vast majority of participants fall into the intended 45-64 age range, with the median participant between 50 and 54 years of age.

---

<sup>4</sup> This chapter was prepared using PRHHP Research Materials obtained from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the PRHHP or the NHLBI.

Table 5.1: Mean Height of Puerto Rican Men by Birth Year

Birth Years	(A) Thieme (1959)		(B) PRHHP		(C) PREHCO	
	Mean	N	Mean	N	Mean	N
1886-90	-	-	164.5	13	-	-
1891-95	164.9	14	163.1	68	-	-
1896-00	163.2	46	164.0	600	162.6	1
1901-05	164.4	71	163.8	1742	155.2	5
1906-10	164.4	193	164.6	2214	162.5	21
1911-15	165.2	197	165.3	2674	161.1	76
1916-20	165.5	252	165.5	2246	162.5	165
1921-25	164.3	380	166.1	297	164.2	248
1926-30	163.8	363	165.7	51	165.8	336
1931-35	163.5	2	-	-	166.7	417
1936-40	-	-	-	-	167.3	482
1941-45	-	-	-	-	166.6	154

Notes: Height is reported in centimeters. The data were collected in 1948-49 for Thieme (1959), in 1965 for the PRHHP, and in 2002-03 for the PREHCO. Note that cohorts born in 1921-25 and earlier in the PREHCO contain men older than age 80 at the time of examination, the age around which most age-related shrinkage commences (Fernihough and McGovern, 2015; Peter et al., 2014). Their advanced age may help to explain why these cohorts are much shorter than in the other two samples.

### 5.3.2 Puerto Rican Elderly: Health Conditions (PREHCO) Project

The second source of data is the first wave of the Puerto Rican Elderly: Health Conditions (PREHCO) Project, an island-wide, cross-sectional sample survey of targeted individuals aged 60 years and up living in Puerto Rico in 2002 and 2003 and their spouses (Palloni et al., 2013). The resident populations of the island municipalities of Culebra and Vieques were excluded from the sample, and the survey includes oversamples of regions heavily populated with Puerto Ricans of African descent and individuals over age 80. Sampling weights post-stratified by age group, sex, and marital status allow estimations to be representative of the population aged 60 years and older in Puerto Rico. A joint team of researchers from the University of Wisconsin-Madison and the Graduate School of Public Health at the University of Puerto Rico collected information from 4,291 respondents and 1,442 spouses through face-to-face interviews which included specialized anthropometric measurements. Only 4.7 percent of those invited refused to participate, and the overall response rate was 93.9 percent. The data set includes a valid height measurement for 89

percent of male participants at least 60 years old. Height is reported to the nearest quarter inch, but there is a high degree of bunching, with 81 percent of measurements reported as whole inches. Although there is a longitudinal component to this survey, with the second wave conducted in 2006 and 2007, I once again only use data from the first wave.

Analysis of the PREHCO data is informative, but just five percent (or 103 men) of the sample was born between 1896 and 1915, and all of these men were 86 years of age or older. As will be discussed in greater detail below, most age-related shrinkage in men occurs after age 80 (Fernihough and McGovern, 2015; Peter et al., 2014), so there is some uncertainty as to what extent differences in height between cohorts is due to shrinkage. The main advantage of the PREHCO data, other than serving as a robustness check for results derived from the PRHHP, is that the birth years of PREHCO participants cover almost the entire period of US direct rule including the Great Depression.

### **5.3.3 Thieme (1959)**

The last source of data is Thieme (1959), a summary of the biological aspects of an anthropological field study of the island population between June 1948 and January 1949 sponsored by the Social Science Research Center of the University of Puerto Rico. A team from Columbia University selected a sample, 3,562 participants in total, to be representative of the Puerto Rican adult population in order to set up a benchmark of biological characteristics as a point of comparison for the future. The sample is composed largely of individuals examined in health centers maintained by the Department of Health. Participants came to health centers not by invitation but instead to receive health services. Researchers made a special effort to select “average” people but could not exclude those with illnesses because they would have “found very few to examine” (Thieme, 1959, p. 22). Selection of sample subjects to meet the calculated distribution was made on the basis of residence and age. Residence quotas were filled by going to certain municipalities for a given number of each sex. Anyone between 20 and 45 years old was taken at random, but younger and older people also voluntarily got in line to be measured and were generally not excluded. I

exclude a sample of 395 prisoners who were added, in part, to fill out the required number of males from some of the regions because prisoners are a highly selected sample; the study confirms that prisoners are shorter (likely of a lower socioeconomic class) and younger than the population in general. Microdata for this study are unavailable, so I estimate mean height by birth cohort using a table of mean height and sample sizes by age in years from a table of the appendix. Mean height by age is reported in centimeters to two decimal places, but height was measured to the nearest centimeter.

## **5.4 Comparisons by Birth Year**

### **5.4.1 Analysis**

In this section I exploit cross-sectional variation in birth year to identify differences in average height over time. Starting with data from the PRHHP, I run an ordinary least squares regression of height in centimeters on a series of birth cohort dummy variables, with the reference group consisting of all individuals born prior to 1896 and thus would have been between the ages of 70 and 79 at the time of the survey. The coefficient on each cohort represents how the average height differed from that of the reference group in 1965.

Notably, I include no control variables in this model. Godoy et al. (2007) caution against interpreting results without controls and go on to add education level and urban/rural status to their regressions. But including such controls essentially removes observable elements of the development process from the birth cohort dummies. As discussed in Section 5.2, educational attainment and urbanization progressed markedly in the early days of US rule. These trends are themselves an outcome of the development process that we hope to capture in the birth cohort dummies. Including such controls would make sense if we were concerned about self-selection into birth cohorts, e.g. that more educated men self-selected into being born later. Of course this is not possible. Instead, controlling for education and urban/rural status leaves the birth cohorts to account for all other elements of the development process, which is not the purpose of these

exercises.

The results of this simple regression appear in column 1 of Table 5.2, and cross-sectional differences in height can be observed in Figure 5.1. The estimated coefficients show a clear pattern of increasing height over time. Even the cohort born between 1906 and 1910 is statistically significantly taller than those born before 1896 at the .10-level, and subsequent cohorts are significantly taller at the .05- or .01-levels. The exception to this trend is the youngest cohort, born between 1926 and 1930, who are on average 2.4 centimeters taller than the reference group but whose coefficient is significant only at the .10-level, likely due to the fact that it has just 51 members. The largest difference is between the reference group and the cohort born between 1921 and 1925, who stand 2.7 centimeters taller on average.

Results from quantile regressions appear in Table D.1 of the appendix and similarly show that later birth cohorts were taller than the reference group. The distribution of heights in general shifted right. Quantile regressions do not discern any differences in height between later cohorts, but this result is likely due to rounding of measurements to the nearest inch (approximately 2.54 cm.), which masks true variation. Regressions for other quantiles produce nearly identical results but sometimes reveal an upward trend across later cohorts.

Contrary to Godoy et al. (2007), the evidence presented above suggests that the biological standard of living improved in early 20th century Puerto Rico. Even when I expand the reference group to include men born between 1886 and 1900 (ages 65-79) in column 2, as done in the previous paper, the pattern of results holds. The 1921-1925 cohort is 2.1 centimeters taller on average than the reference group that includes men between ages 65 and 79, with nearly 9 of 10 falling between the ages of 65 and 69. The primary difference between the conclusions is due to the fact that I do not include age at the time of examination as an explanatory variable because all measurements were taken in the same year, 1965. Variables based on birth and age therefore contain the same information (though in continuous form), so including them both results in severe multicollinearity. Including age adds little information to the regression, inflates the standard errors and diminishes the magnitudes of the birth cohort coefficients, and leads to the erroneous conclusion that there

Table 5.2: Changes in Height Across Birth Cohorts in the PRHP and PREHCO

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Born 1886-00	0.69 (0.78)		0.41 (0.77)	0.38 (0.78)	0.27 (0.78)		
Born 1901-05	0.46 (0.75)	-0.15 (0.27)	0.03 (0.74)	0.16 (0.76)	-0.03 (0.76)		
Born 1906-10	1.27* (0.75)	0.66** (0.27)	0.62 (0.74)	0.94 (0.76)	0.71 (0.76)		
Born 1911-15	1.92** (0.75)	1.31*** (0.26)	0.97 (0.74)	1.52** (0.75)	1.25* (0.75)		
Born 1916-20	2.18*** (0.75)	1.57*** (0.27)	0.94 (0.74)	1.74** (0.76)	1.45* (0.76)	1.36 (1.36)	1.25 (1.34)
Born 1921-25	2.72*** (0.82)	2.12*** (0.42)	1.65** (0.81)	2.52*** (0.82)	2.12*** (0.82)	3.06*** (1.08)	2.76** (1.09)
Born 1926-30	2.35* (1.36)	1.75 (1.16)	1.04 (1.30)	2.30* (1.34)	1.63 (1.37)	4.58*** (1.00)	3.83*** (1.03)
Born 1931-35						5.50*** (0.95)	4.54*** (0.97)
Born 1936-40						6.16*** (0.96)	5.28*** (0.99)
Born 1941-42						5.44*** (1.07)	4.45*** (1.09)
Constant	163.34*** (0.74)	163.95*** (0.23)	166.69*** (0.76)	164.16*** (0.75)	164.57*** (0.75)	161.18*** (0.86)	160.11*** (1.38)
Survey	PRHP	PRHP	PRHP	PRHP	PRHP	PREHCO	PREHCO
Observations	9,805	9,805	9,805	9,805	9,739	1,905	1,869
Controls	None	None	Educ. FE	Urban/Rural	Unskilled	None	Educ. FE
R-squared	0.01	0.01	0.04	0.02	0.02	0.05	0.08

Notes: The dependent variable is standing height in centimeters. Data include men born between 1886 and 1930 for the PRHP and between 1896 and 1942 in the PREHCO. The data were collected in 1965 for the PRHP and in 2002-03 for the PREHCO. Column 2 includes the cohort born between 1896 and 1900 in the reference group for ease of comparison to Godoy et al. (2007). Robust standard errors are shown in parentheses. Significance levels are denoted by \* p < .10, \*\* p < .05, \*\*\* p < .01.

was no secular growth in height. Age and all birth cohort dummies are statistically insignificant when age is included as an explanatory variable (results not shown).

What explains the rise in height? In columns 3 through 5 I add several variables to the regression not as controls, *per se*, but rather to determine the extent to which changes in height are associated with changes in educational attainment, rural-urban migration, and occupations. Educational attainment fixed effects, urban/rural status, and working in an unskilled occupation all decrease the magnitude of the birth cohort dummies, and educational attainment has the largest effect by far. The coefficients on the later birth cohorts remain positive and statistically significant. The takeaway from this exercise is that much of the increase in height in the early 20th century is captured by increases in educational attainment and to a lesser extent by rural-urban migration and a transition towards a more skill-based economy. The direction of causality in these statistical relationships, however, is ambiguous: the urban premium, for example, may be a product of positive selection of internal migrants to San Juan rather than healthier living conditions in the city.

In column 6 of Table 5.2 I use data from the PREHCO and once again regress height in centimeters on a series of birth cohort dummies, with men born between 1896 and 1915 (ages 86-107) as the reference group. The pattern of results closely parallels that derived from the PRHHP. All cohorts are substantially and increasingly taller than the reference group. Birth cohorts from 1921 and later are all statistically significantly taller than the reference group at the .01-level. Those born in 1931 or later stand more than 5 cm. taller than the reference group. Differences between successive cohorts are similarly large: the 1926-30 cohort stands 3.2 cm. taller than the 1916-20 cohort ( $p = .00$ ). Thus, the results from the PREHCO sample also strongly support the conclusion that average height increased in Puerto Rico in the early 20th century. Finally, in column 7 I evaluate the extent to which increases in height over time are captured by greater educational attainment by including educational attainment fixed effects. Once again, the birth cohort coefficients shrink but are still quite large and statistically significant. Education evidently played an important role in increasing stature but was not the only important factor.

Results from quantile regressions using the PREHCO data, found in Table D.1 of the

appendix, once again show that the entire distribution of height shifted to the right over time. In contrast to the PRHHP results, the regressions indicate that growth may have been higher in the upper half of the distribution, suggesting some growth in inequality, although this result varies with the quantiles considered. Lastly, the finer measurements in the PREHCO help to show variation in height between later birth cohorts and confirm that there was a general upward trend in height.

Perhaps the most important and surprising result derived from the PREHCO data is that height continued to increase throughout the Great Depression. The received wisdom is that the island suffered terribly, as “[p]roduction in most of the major branches of the economy fell, incomes contracted drastically, a very large proportion of the labor force was unemployed, and most families of the island experienced great distress” (Perloff, 1952, p. 53). Although height could have continued to grow without a corresponding increase in income, rising stature does not align with claims that “the economy plunged into absolute despondence” in the 1930s (Baer, 1959, p. 450). Improvements in height are in line with GDP estimates by Devereux (2019), who finds that income per capita grew consistently and that Puerto Rico fared better during the Great Depression than the mainland or other Latin economies. Average stature may also have improved in part from the roll out of public health units (PHU) across all municipalities between 1926 and 1937, since the most important proximate source of increasing height in Europe in the last two centuries was the improving disease environment (Hatton, 2014). PHUs drastically expanded public health services and specifically targeted child hygiene, catalyzing unprecedented reductions in mortality (see Chapter 2).

Lastly, in the absence of microdata I consider cross-sectional variation from the 1948-49 study described in Thieme (1959) by simply comparing means across birth cohorts in Table 5.1. For cohorts born between 1896 and 1920, height is once again positively associated with year of birth. Between the 1896-00 and 1916-20 cohorts, average height increases by 2.3 cm. However, in contrast to the PRHHP and the PREHCO, average height falls after 1920. There is a huge, sudden drop-off in stature for the 1921-25 and 1926-30 birth cohorts in the 1948-49 sample. The small stature of these cohorts is anomalous and appears somewhat suspect given that mean height in the

1948-49 sample is statistically indistinguishable from that in the PRHHP for all other cohorts.<sup>5</sup> Consider, too, that the height of these cohorts is substantially lower than in all three other studies of Puerto Ricans listed in Table 5.4, including the PREHCO sample of men ages 71-82 at the time of examination. The reliability of the mean height figures for these cohorts is dubious, but, overall, comparisons across birth cohorts within the 1948-49 sample are less conclusive.

#### 5.4.2 Potential Biases

There are several potential biases that must be considered before drawing conclusions from the cross-sectional analysis. The most serious concern is that earlier birth cohorts are shorter merely because their height has decreased in later life. Recent studies by Peter et al. (2014) and Fernihough and McGovern (2015), however, help to assuage this concern. The former study tracks 85,396 men from a population-based risk factor surveillance program in Austria over 20 years and finds that male height decreases by 1.02 percent on average from age 20 to 85 and that height decline accelerates with age. The latter study uses data from the English Longitudinal Study of Aging, which follows 3,098 males over age 50 over the course of a decade. They estimate a considerably larger loss in height, around 2-4 cm. over the life course, but a high proportion of all shrinking occurs after age 80.<sup>6</sup>

Taking into account shrinkage, there is still evidence of a rise in the height of Puerto Ricans in the PRHHP data. The 1921-25 cohort (ages 40-44) is 0.54 cm. taller than the 1916-20 cohort (ages 45-49), though the difference is not quite statistically significant at conventional levels ( $p = .15$ ). The difference is large since we should not expect any discernible amount of shrinking before age 50 and the difference in birth years is so minor. Perhaps more striking is that the 1921-25 cohort stands 0.80 cm. taller on average than the 1911-15 cohort (ages 50-54) and 1.45 cm. taller than the 1906-10 cohort (ages 55-59)—significant at the .05- and .01-levels, respectively. Additionally,

---

<sup>5</sup> The mean height of cohorts born before 1921 in the PREHCO are all much shorter than the corresponding cohorts in Thieme (1959), but this is likely due to age-related shrinkage after age 80, which I discuss in greater detail in the following subsection.

<sup>6</sup> For instance, estimates based off of a decade between measurements suggest that the average 70-year-old in their sample has lost less than 1 cm. off their lifetime maximum height, the average 80-year-old has lost about 2 cm., and the average 90-year-old has lost more than 3 cm (Fig. 4).

the point estimates for all birth cohorts from 1911 through 1930 are larger than a conservative benchmark of 1.87, the expected loss by age 85, based on the analysis by Peter et al. (2014), for the tallest man in the PRHHP. Plus the 1921-1925 cohort stands 2.7 centimeters taller on average than the reference group—a group with an average age of just 72.3, making the 1.87 cm. reference even more conservative.

Age-related shrinkage in the PREHCO sample is much more of a concern, since the reference group is comprised of men ages 86 to 107. Taking into consideration shrinkage, the PREHCO results still suggest an increase in height, however, as the four youngest cohorts are all more than 4 cm. taller than the reference group. There is also a general rise in height across successive cohorts. The difference in average height between men born in 1926-30 (ages 71-77) and 1936-40 (ages 61-67), for example, is 1.58 cm., significant at the .01-level.

In contrast, two sources of bias could possibly lead to an understatement of the increase in stature over time. First, the first wave of the PRHHP was carried out in 1965, near the end of the Puerto Rican Great Migration. From about 1946 to 1965, hundreds of thousands of Puerto Ricans emigrated to the mainland such that by 1970 one-fourth of those born in Puerto Rico lived outside of the island (Ruggles et al., 2018). Relative to non-migrants in Puerto Rico, migrants to the US tended to be more highly skilled, more educated, and overwhelmingly younger (Castillo-Freeman and Freeman, 1992; Sandis, 1970). Thieme (1957) finds evidence suggesting that migrants tended to come from a better socioeconomic background which is expressed through diet and health in larger body dimensions. Hence, the youngest cohorts in the PRHHP might have lost a disproportionate amount of their taller members to migration, leading to a negative bias of their coefficients in Table 5.2.

Unfortunately, there is not a data set containing information on the height of Puerto Rican migrants in the US suitable for comparison to the PRHHP, so the degree of bias introduced by selective migration is unknown.<sup>7</sup> However, in Table D.2 of the appendix, I pool the one-percent

---

<sup>7</sup> The Hispanic Health and Nutrition Examination Survey (HHANES) (1985), conducted from 1982 to 1984, includes 101 men born in Puerto Rico between 1910 and 1930 and residing in the New York City area in 1984. Mean height of this group is 165.2 cm., which is slightly shorter than the PRHHP for these birth years. However,

samples of the US and Puerto Rican censuses from 1970 to compare predicted height by cohort with and without migrants included in the sample. Specifically, I use coefficients estimated from a regression of height on age cohort and educational attainment in the PRHHP and make out-of-sample predictions of the height of men born in Puerto Rico in the census. The analysis is imperfect, in part because the coefficients on the birth cohort dummies may still be biased if migration selectivity is not perfectly captured by educational attainment. Moreover, the US census samples for the earliest birth cohorts are quite small, with just 11, 15, 45, and 60 observations, respectively. Still, the evidence in Table D.2 suggests that bias from selective migration is unimportant: the coefficient estimates in for the full sample (column 1) and for non-migrants only (col. 2) are indistinguishable, and there was a less pronounced upward trend in predicted height across migrant birth cohorts (col. 3).

Second, survivorship bias might have increased the average height of earlier birth cohorts in the PRHHP and the PREHCO. Although stature may negatively affect longevity all else equal (Samaras, 2012), previous research has found a positive correlation between height and life expectancy, especially in developing countries, due to better nutrition and access to resources (Akachi and Canning, 2015; Costa, 2007; Deaton, 2007; Harris, 2007). Earlier birth cohorts should have experienced higher mortality rates before the surveys in 1965 and 2002-03, and those who died were likely shorter on average. Hence, the deaths of shorter men in the earlier birth cohorts might lead to an understatement of the rise in height.

Although an evaluation of mortality between the several waves of surveys cannot determine the extent of selective mortality for those deceased prior to the surveys, it is informative that there is a weak negative relationship between height and mortality in both the PRHHP and PREHCO. A linear probability model with controls for birth quinquennium finds that a one-centimeter increase in height is associated with a 0.07 percentage point decrease in the probability of death between the first wave of the PRHHP in 1965 and the fourth wave in 1980, but the relationship is not statistically

---

the comparison is not of much value since the flow of migrants to the US had turned negative and a large wave of positive return migration had commenced in the two decades between the PRHHP and HHANES (Castillo-Freeman and Freeman, 1992; Ramos, 1992).

( $p = 0.28$ ). Likewise, a one centimeter increase in height is associated with a 0.13 percentage point decrease in the probability of death between the first wave of the PREHCO in 2002-03 and the second wave in 2006-07, but the relationship is also statistically insignificant ( $p = 0.39$ ).

Lastly, an alternative explanation for the results above is that selection into the sample varies across birth cohorts. This is a concern because the changes in observed heights over time may be driven by changes in the probability that individuals of different heights were differentially selected into the sample (Bodenhorn et al., 2014).

There are two reasons, however, to doubt that selection bias is driving the results. First, in the case of the PRHHP, all men ages 45 to 64 in the enumerated districts were invited to participate, and the vast majority accepted: 79.9 percent of urban men and 82.4 percent of rural men (Garcia-Palmieri et al., 1984). Likewise, only 4.7 percent of those invited to the PREHCO refused to participate, and the overall response rate was 93.9 percent. The response rate in Thieme (1959) is less certain, but the study reports that anyone between 20 and 45 years old was taken at random. At a minimum, there was no selection by design in any of the three surveys.

Second, data from the 1970 census of Puerto Rico exhibit similar patterns of consistent improvements in educational attainment, a key predictor of height, as the PRHHP and PREHCO data (Table D.3 of the appendix). In each data set there is a general upward trend in educational attainment, starting with less than four years in the late 19th century, rising to around eight years by 1930, and reaching more than nine years by 1940. Mean years of education within birth cohorts and across data sets are quite comparable as well. Therefore, the census data offer no indication that the results are driven by selection bias.

## 5.5 Comparisons by Age

In the previous section I used cross-sectional variation in birth year to show that there was an increase in the height of Puerto Rican males over time, and I argued that the differences are not attributable to age-related shrinkage. Still, comparisons by age groups over time serve as an important robustness check and may be preferable for estimating a more precise magnitude of

changes in height over time. In this section, I compare overlapping age groups in the three surveys to assess changes in height without concerns about shrinking.

To make such comparisons, I assume that differences in height between surveys are not the product of differences in how height was measured. As previously mentioned, height was recorded to the nearest inch (approximately 2.54 cm.) in the PRHHP, to the nearest quarter inch (0.64 cm.) in the PREHCO, and the nearest centimeter in Thieme (1959). Such differences should not distort comparisons across surveys, however: in the PREHCO, for example, rounding measurements to the nearest inch changes cohort sample means by about 0.1 to 0.2 cm.—not nearly enough to materially affect the conclusions drawn from the analysis. In the PRHHP, heights shorter than 58 inches are bottom-coded, and those taller than 72 inches are top-coded. Once again, adjusting PREHCO measurements in a similar fashion changes cohort sample means by 0.04 cm. on average and by 0.15 cm. at most. I nonetheless top- and bottom-code the PREHCO data in Table 5.3 and Figure 5.1 to improve comparisons to the PRHHP. In all three surveys, standing height was taken without shoes. PREHCO measurements were taken with a stadiometer, while the instruments used to take measurements in other two surveys are unknown. However, it is reasonable to assume that height measurements across surveys are comparable since there is little disagreement about the definition of height. For instance, there were no statistically significant differences in the means of height measurements of matched samples taken by two different anthropometrists in Thieme (1959), despite significant differences in fourteen other anthropometric measurements.

The mean heights of Puerto Rican men by age at the time of examination appear in Table 5.3, with the surveys listed in chronological order from left to right, and Figure 5.1. A comparison of data collected in 1948-49 for Thieme (1959) and data collected in 1965 for the PRHHP suggests that there was a rise in average height over time. For both age groups for which there are fairly large samples—ages 40-44 and 45-49—the increase in height is sizable at around one and a half cm., and the difference is statistically significant at the .10- and .01-levels, respectively. The magnitude of the increase is quite large at around one cm. for ages 35-39 and 50-54, too, but the samples are too small for inference. For one cohort, ages 55-59, the Thieme (1959) sample is taller by 0.3 cm.,

but once again the group contains too few measurements for inference.

Table 5.3: Mean Height of Puerto Rican Men by Age

Ages	(A) Thieme (1959)		(B) PRHHP		(B) PREHCO	
	Mean	N	Mean	N	Mean	N
80-114	-	-	-	-	163.0	343
75-79	-	-	164.5	13	164.6	285
70-74	-	-	163.1	68	166.6	373
65-69	-	-	164.0	600	166.6	450
60-64	-	-	163.8	1742	167.6	454
55-59	164.9	14	164.6	2114	-	-
50-54	164.2	18	165.3	2674	-	-
45-49	163.9	99	165.5	2246	-	-
40-44	164.7	122	166.1	297	-	-
35-39	164.8	179	165.7	51	-	-
30-34	165.3	227	-	-	-	-
25-29	164.8	324	-	-	-	-
20-24	163.9	475	-	-	-	-
17-19	163.7	60	-	-	-	-

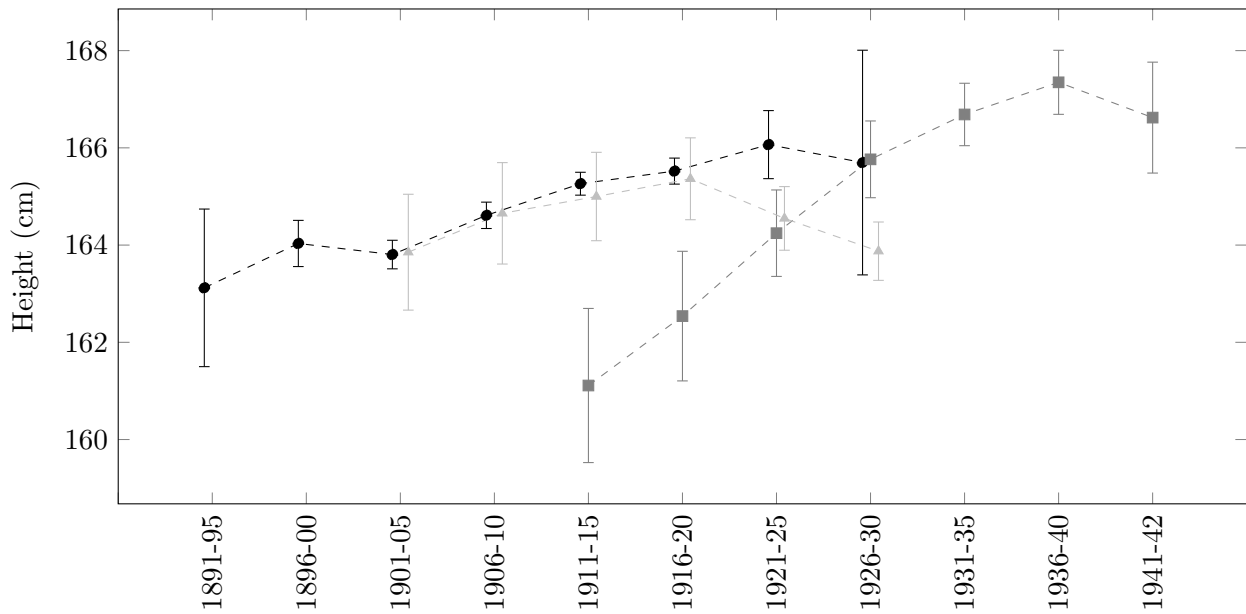
Notes: Height is reported in centimeters. The data were collected in 1948-49 for Thieme (1959), in 1965 for the PRHHP, and in 2002-03 for the PREHCO. PREHCO measurements have been rounded and bottom- and top-coded to improve comparisons to the PRHHP.

A comparison of the 1965 PRHHP and 2002-03 PREHCO data likewise suggests that there was an increase in height over time. The cohorts ages 60-64, 65-69, and 70-74 are 3.8, 2.6, and 3.5 cm. taller in the PREHCO than in the PRHHP, and each difference is significant at the .01-level. The last overlapping cohort, ages 75-79, is 0.1 cm. taller in the PREHCO sample, but this group only includes 13 PRHHP participants.

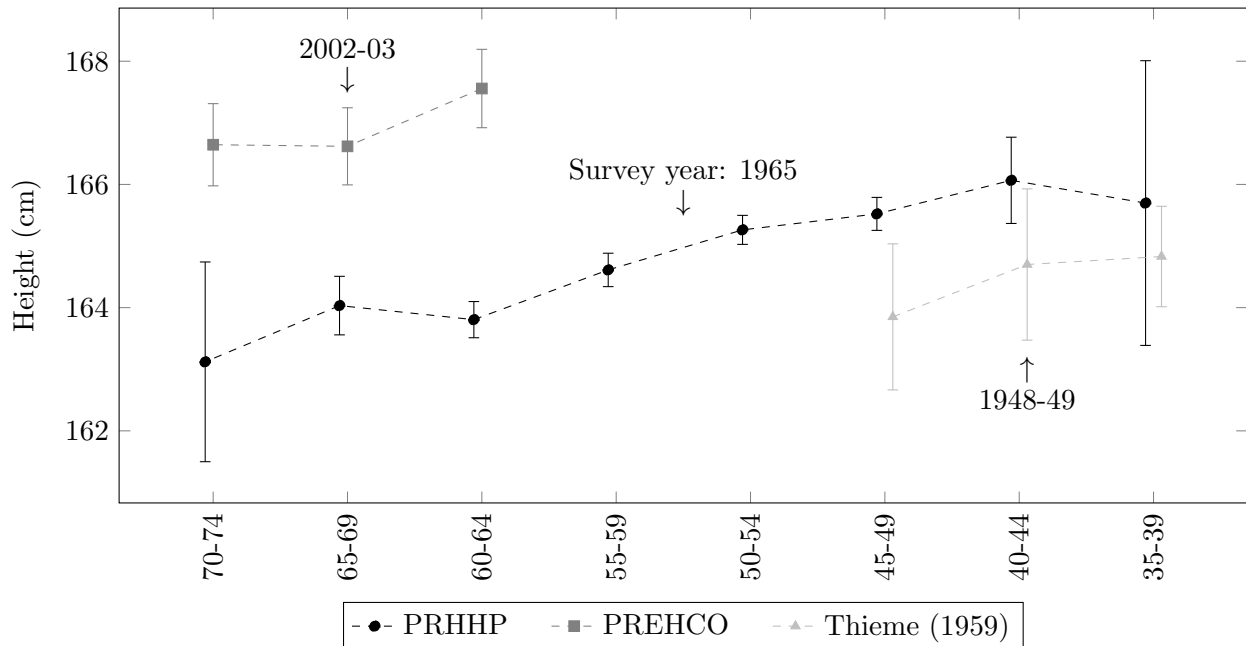
Taken together, these results strongly suggest a large increase in height over the span of a half century. Controlling for age cohort, participants in the PRHHP are 1.3 cm. taller on average than those in Thieme (1959), while participants in the PREHCO are 3.3 cm. taller on average than those in the PRHHP. These increases equate to growth rates of 0.81 and 0.87 cm./decade, respectively. These comparisons include birth cohorts from 1886 to 1942, though with much less coverage in the earliest years. A back-of-the-envelope calculation combining these two differences says that height increased by about 0.84 cm./decade, or 4.2 cm. in total, from roughly 1890 to 1940.

Figure 5.1: Mean Height of Puerto Rican Men by Birth Year and Age

(a) Mean Height by Birth Cohort



(b) Mean Height by Age Cohort



Notes: Error bars denote 95% confidence intervals. Cohorts with fewer than 50 observations are excluded. PREHCO measurements have been rounded and bottom- and top-coded to improve comparisons to the PRHHP. For more information, see Tables 5.1 and 5.3.

This estimate is reasonable, if not conservative, when we consider that the US Army found that its average recruit from Puerto Rico in WWII stood two inches (5.1 cm.) taller than its average recruit in WWI just over two decades earlier (Chamber of Commerce of Puerto Rico, 1945). Given the various sources of bias in the cross-sectional comparisons, including imprecision regarding the extent of shrinkage, I consider 4.2 cm. to be my preferred estimate of the growth in stature from late Spanish rule to essentially the end of US direct rule.<sup>8</sup>

## 5.6 Cross-Country Comparison

That Puerto Ricans experienced improvements in living conditions is not surprising when considered alongside data from countries with shared ethnic, cultural, and historical backgrounds. By midcentury, Puerto Rico trailed only Argentina and Uruguay within Latin America in terms of GDP per capita and lagged behind only Uruguay in terms of consumption per capita (Devereux, 2019). According to the Chamber of Commerce of Puerto Rico,

Puerto Rico's citizens own more real estate, better homes, more automobiles, more radios, electric refrigerators, furniture, clothing, and machinery than the citizens of any other Caribbean country. They consume more medicines, have more hospitals, get better medical and dental treatment and use more soap than the people of any other Caribbean area. (1945, p. 2)

Puerto Rico was also the most intensively electrified area south of the Tropic of Cancer and exported manufactured goods more than twice the total value of its nearest competitor in the Caribbean. Most relevant to this inquiry, Puerto Rico boasted the second lowest infant mortality rate (behind Uruguay) and the third highest rate of primary school enrollment (behind Uruguay and Argentina) in Latin America at mid-century (Molina and Noam, 1964; Nuñez, 2005). The island also experienced the highest rate of increase in literacy in all of Latin America and the Caribbean

---

<sup>8</sup> Note that the results from the analysis of the PRHHP data suggest a comparable rate of growth (0.97 cm./decade).

from 1900 to 1950, and its progress in public health from 1930 to 1960—in addition to improvements in the immediate aftermath of US annexation—had “no parallel in any other country” according to a former commissioner of health (Bourne and Bourne, 1966, p. 81). These indicators suggest that Puerto Ricans ought to have been relatively tall for their region, since exposure to disease is one of the most important non-genetic factors determining height, and height has been found to be positively associated with education (Hatton, 2014; Silventoinen, 2003).

Were Puerto Ricans also taller than citizens of other countries in the region? Direct comparisons across studies must be carried out with caution because of varying degrees of selection in samples. Still, it is clear from Table 5.4 that Puerto Ricans compared quite favorably to other Latin Americans in terms of height despite their sizable gap with the mainland US.<sup>9</sup> Puerto Ricans born in the 1920s were among the tallest Latin Americans outside of Argentina and Uruguay and may have been taller than Spaniards.

Cuba serves as a particularly interesting comparison to Puerto Rico in the early 20th century for its proximity and shared cultural and political history. Cuba and Puerto Rico were the last of Spain’s New World colonies, both breaking ties with the empire in 1898. Like Puerto Rico, Cuba’s economy in the early 20th century was dominated by cane sugar and other commercial agriculture with a high degree of control by Americans. Before its communist revolution, Cuba was remarkably prosperous; in the 1920s, Cuban GDP per capita was roughly as high as some states in the US South and was within striking distance of Western Europe (Ward and Devereux, 2012). Surveying the Cuban anthropometric evidence over more than five decades, Domínguez (1978, p. 515) concludes that the height of both sexes and three racial categories (white, black, and mixed-race) “increased substantially during the 20th century.” It would therefore be surprising if Puerto Rico did *not* experience an increase in average height over this period.

---

<sup>9</sup> The table excludes nearby Caribbean nations with populations predominately of African origin due to the well-documented large stature of Africans in spite of relative health deficiencies. Puerto Ricans in this era were roughly as tall as Haitians (Basu et al., 1976) but shorter than Jamaicans and Trinidadians (Ashcroft et al., 1966, US Center for Chronic Disease Control, 1967). None of the three data sets used in this paper include information on race, although the third wave of the PREHCO does. I do not present results based on the third wave of the PREHCO because there is a high rate of attrition between the waves. In any case, results do not meaningfully change when using that data and excluding Puerto Ricans who identify as black or African-American.

Table 5.4: Mean Male Height in Latin America, Spain, and the United States

Country	Height (cm)	Birth	Age	Year	N	Source
Argentina <sup>†</sup> <sup>^</sup>	169.1	1921-30	18	1939-48	≈1,000	Salvatore (2004)
Bolivia	161.9	-	-	1962	2,371	US Center for Chronic Disease Control (1967)
Chile <sup>†</sup>	165.0	-	-	1960	4,267	US Center for Chronic Disease Control (1967)
Colombia	164.5	1920-29	18-21	1938-50	≈25,000	Meisel and Vega (2007)
Cuba	168.3	1943-46	18.5-20.5	1963-64	164	Laska-Mierzejewska (1970)
Ecuador <sup>†</sup>	161.0	-	-	1959	2,438	US Center for Chronic Disease Control (1967)
Guatemala <sup>†</sup>	154.8	-	-	1959-64	158	Russell (1976)
Mexico <sup>†</sup>	≈164	1920-29	18-45	1938-74	468	López-Alonso and Porras Condey (2003)
Paraguay <sup>†</sup>	167.2	1917-47	18-48	1965	523	US Center for Chronic Disease Control (1967)
Peru <sup>†</sup>	161.0	-	-	1959	6,527	US Center for Chronic Disease Control (1967)
Puerto Rico	167.6	1927-47	19-39	1966	60	Fernández et al. (1971)
Puerto Rico	166.0	1921-30	35-44	1965	348	García-Palmieri et al. (2002)
Puerto Rico	165.1	1921-30	71-82	2002-03	584	Palloni et al. (2013)
Puerto Rico	164.1	1921-30	18-27	1948-49	743	Thieme (1959)
Spain	164.8	1921-30	18-21	1939-51	-	Martinez Carrión and Pérez Castejón (1998)
Spain	167.8	1923-30	56-79	1986-2006	≈4,000	Spijker et al. (2012)
United States	175.3	1921-30	41-54	1971-75	827	Godoy et al. (2007)
Uruguay <sup>†</sup>	168.7	-	-	1962	569	US Center for Chronic Disease Control (1967)
Venezuela	164.0	1914-43	20-49	1963	231	US Committee on Nutrition (1964)

Birth years and ages have been imputed where required. The Cuban and US samples are limited to white males.

<sup>†</sup> Military and/or police data.

<sup>‡</sup> Rural sample.

<sup>^</sup> Unskilled, urban but not “big city” sample.

At first glance, the data in Table 5.4 suggest that Cubans were significantly taller than Puerto Ricans. However, all survey participants in Laska-Mierzejewska (1970) resided in Cuba's most important metropolitan area, Havana. Schools were chosen for investigation "in such a way as to maintain a cross-section of all social and economic classes of the population" (p. 581), but the sample nonetheless likely represents an above average standard of living. Plus, if we compare the two youngest cohorts of the PRHHP living in urban areas (N=201), the mean height is 167.1 centimeters. This point estimate is of course lower than the corresponding Cuban figure, but the Cubans would have been born around a decade and a half later. The mean height of Puerto Ricans born in the early 1940s (ages 60-62) from the PREHCO sample is 167.2 cm. (N=636). In 1966, an island-wide, representative sample (N=60) of the Puerto Rican population found an average height of 167.6 centimeters for men between the ages of 19 and 39 (or born between 1927 and 1947) (Fernández et al., 1971). Importantly, half of that sample came from rural areas. Along with the results of the PRHHP study, we can conclude with some confidence that the height of Puerto Ricans was comparable to that of Cubans around mid-century.<sup>10</sup>

## 5.7 Discussion and Conclusion

The dominant view in the historical literature holds that the living conditions of most Puerto Ricans did not improve or were made worse during the period of direct rule by the US, roughly from the end of the Spanish-American War to the end of WWII. The anthropometric evidence presented in this paper, however, suggests that living conditions in Puerto Rico improved in the decades immediately following US annexation. This is not to say that the transfer of power from Spain to the US had a causal impact on the increase in height or that the observed increase was larger than would have occurred under an alternative political arrangement. Nonetheless, this result contributes to a mounting body of quantitative evidence to overturn the historical narrative

---

<sup>10</sup> We may also consider Knott (1963), which finds a mean height of 138.3 centimeters for 73 private school boys in San Juan in 1962. This average is clearly higher than the average height of 136.5 centimeters for 10.5-year-olds in Havana in 1963 (Laska-Mierzejewska, 1970). However, given the uncertain degree of selection in these samples, as well as the fact that 10-year-olds in San Juan would have already reaped benefits from post-WWII industrialization—a period of growth agreed upon by scholars throughout the literature—this comparison is much less informative.

that early US rule impoverished Puerto Rico. Living conditions in the early 20th century were undoubtedly low by US or Western European standards, but they were a marked improvement over the conditions that prevailed under Spanish rule.

Not only did the height of Puerto Rican men rise after US annexation, the increase of 4.2 cm. over a half century, or about 0.84 cm./decade, is large for this period. Growth in stature was highest in Western Europe and its offshoots, as average adult height by birth cohort increased by about 1.08 cm./decade from 1891-95 to 1941-45 in 11 Western European countries (Hatton and Bray, 2010). Although slower than the Western European average, Puerto Rico outpaced France, Spain, and Sweden. Average height increased by only 1.9 cm., or 0.38 cm./decade, in Latin America and the Caribbean from the 1890s to the 1940s (Baten and Blum, 2014). In other words, male height in Puerto Rico increased at a rate more than double the regional average.<sup>11</sup>

An increase of 4.2 cm. over a half century is also economically significant. Using data on the height of women from 1951 to 1992, Akachi and Canning (2015) estimate that each centimeter gain within countries is associated with a six percent increase in income per capita. Similarly, Baten and Blum (2012) find that an additional centimeter in stature was associated with more than a ten percent increase in GDP between 1870 to 1949. The increase in height described in this paper therefore translates to major improvements in economic conditions and complements the recent estimate by Devereux (2019) that income per capita in 1940 was 2.7 times higher than in 1900.

The roots of Puerto Rico's impressive socioeconomic transformation in the 20th century began, contrary to most accounts, before WWII. Evidence of improvements in height presented in this paper helps to correct a fundamental inaccuracy in the literature on the economic history of Puerto Rico, motivating a reconsideration of US rule and opening up opportunities for research on US interventions in public health, education, infrastructure, and other areas of society.

---

<sup>11</sup> The NCD Risk Factor Collaboration (NCD-RisC) (2016), which includes data from over 2,545 population-based surveys, ranks Puerto Rico 13th among 200 countries and territories, and 2nd in the Western Hemisphere, in terms of the increase in average male height from the 1896 to the 1945 birth cohort. However, there is little information available about the sources of data, and estimates of increases for Puerto Rico and other countries in the region are substantially higher than in other studies. For example, the estimated increase in mean male height in Puerto Rico between 1896 and 1945 is 8.97 cm.

## Bibliography

- Abraham, S. and Sun, L. (forthcoming). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics.
- Akachi, Y. and Canning, D. (2015). Inferring the economic standard of living and health from cohort height: Evidence from modern populations in developing countries. Economics and Human Biology, 19:114–128.
- Alsan, M. and Goldin, C. (2019). Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880 to 1920. Journal of Political Economy, 127(2):586–638.
- Anderson, D. M., Brown, R., Charles, K. K., and Rees, D. I. (2020a). Occupational licensing and maternal health: Evidence from early midwifery laws. Journal of Political Economy, 128(11).
- Anderson, D. M., Charles, K. K., Las Heras Olivares, C., and Rees, D. I. (2019a). Was the first public health campaign successful? American Economic Journal: Applied Economics, 11(2):143–175.
- Anderson, D. M., Charles, K. K., and Rees, D. I. (forthcoming). Re-examining the contribution of public health efforts to the decline in urban mortality. American Economic Journal: Applied Economics.
- Anderson, D. M., Charles, K. K., Rees, D. I., and Wang, T. (2019b). Water Purification Efforts and the Black-White Infant Mortality Gap, 1906-1938. (NBER Working Paper 26489). Cambridge, MA: National Bureau of Economic Research.
- Anderson, D. M., Rees, D. I., and Wang, T. (2020b). The phenomenon of summer diarrhea and its waning, 1910-1930. Explorations in Economic History, 78.
- Arbona, A. (1942). Malaria control through the public health units. Puerto Rico Health Bulletin, 6(9):254–263.
- Arbona, G. (1937). School hygiene in Puerto Rico. Puerto Rico Health Bulletin, 1(9):195–199.
- Arbona, G. (1940). Infant hygiene in Puerto Rico. Puerto Rico Health Bulletin, 4(8):229–231.
- Ashcroft, M., Ling, J., Lovell, H., and Miall, W. (1966). Heights and weights of adults in rural and urban areas of Jamaica. British Journal of Preventive and Social Medicine, 20(1):22–26.
- Astorga, P., Berges, A. R., and Fitzgerald, V. (2005). The standard of living in Latin America during the twentieth century. Economic History Review, 58(4):765–796.

- Ayala, C. J. and Bergad, L. W. (2002). Rural Puerto Rico in the early twentieth century reconsidered: Land and society, 1899-1915. Latin American Research Review, 37(2):65–97.
- Ayala, C. J. and Bergad, L. W. (2020). Agrarian Puerto Rico: Reconsidering Rural Economy and Society, 1899-1940. New York: Cambridge University Press.
- Ayala, C. J. and Bernabe, R. (2007). Puerto Rico in the American Century: A History Since 1898. Chapel Hill, NC: University of North Carolina Press.
- Ayuda, M. I., Collantes, F., and Pinilla, V. (2010a). From locational fundamentals to increasing returns: The spatial concentration of population in Spain, 1787-2000. Journal of Geographical Systems, 12(1):25–50.
- Ayuda, M. I., Collantes, F., and Pinilla, V. (2010b). Long-run regional population disparities in Europe during modern economic growth: A case study of Spain. Annals of Regional Science, 44(2):273–295.
- Baer, W. (1959). Puerto Rico: An evaluation of a successful development program. Quarterly Journal of Economics, 73(4):645–671.
- Bailey, M., Clay, K., Fishback, P., Haines, M. R., Kantor, S., Severnini, E., and Wentz, A. (2018). ICPSR: Version V2 [dataset]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Data retrieved from <https://doi.org/10.3886/ICPSR36603.v2>.
- Bailey, M. J. and Goodman-Bacon, A. (2015). The War on Poverty’s experiment in public medicine: Community health centers and the mortality of older Americans. American Economic Review, 105(3):1067–1104.
- Banerjee, A., Duflo, E., and Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in china. Journal of Development Economics, 145.
- Bary, H. V. (1923). Child Welfare in the Insular Possessions of the United States: Part I, Porto Rico. Washington, DC: Government Printing Office.
- Basu, A., Namboodiri, K., Weitkamp, L., Brown, W., Pollitzer, W., and Spivey, M. (1976). Morphology, serology, dermatoglyphics, and microevolution of some village populations in Haiti, West Indies. Human Biology, 48(2):245–269.
- Baten, J. and Blum, M. (2012). Growing tall but unequal: New findings and new background evidence on anthropometric welfare in 156 countries, 1810-1989. Economic History of Developing Regions, 27:S66–S85.
- Baten, J. and Blum, M. (2014). Why are you tall while others are short? Agricultural production and other proximate determinants of global height. European Review of Economic History, 18:144–165.
- Bauernschuster, S., Driva, A., and Hornung, E. (2019). Bismarck’s health insurance and the mortality decline. Journal of the European Economic Association, 17(6):1–47.
- Baum-Snow, N., Henderson, J. V., Turner, M. A., and Zhang, Q. (2020). Does investment in national highways help or hurt hinterland city growth? Journal of Urban Economics, 115.

- Baumol, W. J. and Wolff, E. N. (1996). Catching up in the postwar period: Puerto Rico as the fifth 'tiger'? World Development, 24(5):869–885.
- Beeson, P. E., DeJong, D. N., and Troesken, W. (2001). Population growth in US counties, 1840-1990. Regional Science and Urban Economics, 31(6):669–699.
- Berbusse, E. J. (1966). The United States in Puerto Rico: 1898-1900. Chapel Hill, NC: University of North Carolina Press.
- Berg, C. N., Deichmann, U., Liue, Y., and Selod, H. (2017). Transport policies and development. Journal of Development Studies, 53:465–480.
- Berger, T. and Enflo, K. (2017). Locomotives of local growth: The short- and long-term impact of railroads in Sweden. Journal of Urban Economics, 98:124–138.
- Bertazzini, M. C. (forthcoming). The long-term impact of Italian colonial roads in the Horn of Africa, 1935-2000. Journal of Economic Geography.
- Bhalotra, S., Karlsson, M., and Nilsson, T. (2017). Infant health and longevity: Evidence from a historical intervention in Sweden. Journal of the European Economic Association, 15(5):1101–1157.
- Bird, J. and Straub, S. (2020). The Brasília experiment: The heterogenous impact of road access on spatial development in Brazil. World Development, 127:1–13.
- Bleakley, H. (2007). Disease and development: Evidence from hookworm eradication in the American South. Quarterly Journal of Economics, 122(1):73–117.
- Bobonis, G. J. and Morrow, P. M. (2014). Labor coercion and the accumulation of human capital. Journal of Development Economics, 108:32–53.
- Bodenhorn, H., Guinnane, T. W., and Mroz, T. A. (2014). Sample-Selection Bias in the Historical Heights Literature. (Cowles Foundation Working Paper). New Haven, CT: Yale University.
- Borusyak, K. and Jaravel, X. (2018). Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume. (Working Paper).
- Bosker, M., Brakman, S., Garretsen, H., de Jong, H., and Schramm, M. (2008). Ports, plagues and politics: Explaining Italian city growth 1300-1861. European Review of Economic History, 12(1):97–131.
- Bosker, M., Deichmann, U., and Roberts, M. (2018). Hukou and highways: The impact of China's spatial development policies on urbanization and regional inequality. Regional Science and Urban Economics, 71:91–109.
- Boswell, T. D. (1978). Internal migration in Puerto Rico prior to economic development. Social and Economic Studies, 27(4):434–463.
- Bothwell, L. D. (1937). The geographic factor in the distribution, trends, and peaks of the population of Puerto Rico, 1900-1930. Papers of the Michigan Academy of Science, Arts, and Letters, 22:321–327.

- Bourne, D. D. and Bourne, J. R. (1966). Thirty Years of Change in Puerto Rico: A Case Study of Ten Selected Rural Areas. New York: Frederick A. Praeger, Publishers.
- Bravo, A. L. (1958). Development of medical care services in Latin America. American Journal of Public Health, 48(4):434–447.
- Brennan, L., McDonald, J., and Shlomotiz, R. (1997). Toward an anthropometric history of Indians under British rule. Research in Economic History, 17:185–246.
- Brown, P. (2015). Model-based geostatistics the easy way. Journal of Statistical Software, 63(12):1–24.
- Browning, H. L. (1958). Recent trends in Latin American urbanization. Annals of the American Academy of Political and Social Science, 316(1):111–120.
- Buckwalter, L. M. (2019). The impact of a ‘colonizing river’: Colonial railways and the indigenous population in French Algeria at the turn of the century. Economic History of Developing Regions, 34(1):16–47.
- Bulmer-Thomas, V. (2003). The Economic History of Latin America Since Independence (2nd ed.). New York: Cambridge University Press.
- Burdekin, R. C. and Laney, L. O. (2008). Financial market reactions to the overthrow and annexation of the Hawaiian kingdom: Evidence from London, Honolulu, and New York. Cliometrica, 2(2):119–141.
- Bütikofer, A., Løken, K. V., and Salvanes, K. G. (2019). Infant health and long-term outcomes. Review of Economics and Statistics, 101(2):341–354.
- Cabán, P. A. (1999). Constructing a Colonial People: Puerto Rico and the United States, 1898-1932. Boulder, CO: Westview Press.
- Cabán, P. A. (2002). The colonizing mission of the US in Puerto Rico. In A. Sampaio, C. Vélez-Ibañez, and M. González-Estay (Eds.), *Transnational Latino/a Communities: Re-examining Politics, Processes, and Culture* (pp. 115-145). Lanham, MD: Rowman and Littlefield.
- Camaschella, C. (2015). Iron-deficiency anemia. New England Journal of Medicine, 372(19):1832–1843.
- Carroll, H. K. (1899). Report on the Island of Porto Rico. Washington, DC: Government Printing Office.
- Castillo-Freeman, A. and Freeman, R. B. (1992). When the minimum wage really bites: The effect of the U.S.-level minimum wage on Puerto Rico. In G.J. Borjas and R.B. Freeman (Eds.), *Immigration and the Workforce: Economic Consequences for the United States and Source Areas* (pp. 177-212). Chicago, IL: University of Chicago Press.
- Chamber of Commerce of Puerto Rico (1945). A Half Century of Progress: An Appraisal of What Puerto Rico Has Done to Improve the Living Standards, Health, and Productivity of its People Since the Beginning of the Twentieth Century. San Juan, PR: Publications of the Chamber of Commerce of Puerto Rico.

- Chaves, J. (1937a). The public health units in Puerto Rico. Puerto Rico Health Bulletin, 1(2):26–27.
- Chaves, J. (1937b). Shoes, versus other means of controlling uncinariasis. Puerto Rico Health Bulletin, 1(5):88–91.
- Chaves, J. (1938a). Development of public health units and significant accomplishments during fiscal years 1935-36—1936-37—1936-37. Puerto Rico Health Bulletin, 2(9):271–278.
- Chaves, J. (1938b). Public health nursing in Puerto Rico. Puerto Rico Health Bulletin, 2(7):223–227.
- Chaves, J. (1938c). Public health units as a part of the organization of the insular health department. Puerto Rico Health Bulletin, 2(1):43–48.
- Chaves, J. (1942). History, development and functioning of public health units in Puerto Rico. Puerto Rico Health Bulletin, 6(9):221–241.
- Clark, V. S. (1930). Porto Rico and its Problems. Washington, DC: Brookings Institution.
- Clay, K., Juul Egedesø, P., Worm Hansen, C., Sandholt Jensen, P., and Calkins, A. (2020). Controlling tuberculosis? Evidence from the first community-wide health experiment. Journal of Development Economics, 146.
- Clay, K., Schmick, E., and Troesken, W. (2019). The rise and fall of pellagra in the American South. Journal of Economic History, 79(1):32–62.
- Colella, F., Lalive, R., Ocran Sakalli, S., and Thoenig, M. (2019). Inference with Arbitrary Clustering. (IZA Discussion Paper No. 12584). Bonn, Germany: Institute of Labor Economics.
- Conley, T. G. (1999). GMM estimation with cross sectional dependence. Journal of Econometrics, 92(1):1–45.
- Coryllos, P. N. (1937). The antituberculosis organization in Puerto Rico. Puerto Rico Health Bulletin, 1(2):51–60.
- Costa, D. L. (2007). Height, weight, wartime stress, and older age mortality: Evidence from the Union Army records. Explorations in Economic History, 30(4):424–449.
- Costa, D. L. and Kahn, M. E. (2004). Changes in the value of life, 1940-1980. Journal of Risk and Uncertainty, 29(2):159–180.
- Curtis, K. J. and Scarano, F. (2011). Puerto Rico's population *padrones*, 1779-1802. Latin American Research Review, 46(2):200–213.
- Cutler, D. and Miller, G. (2005). The role of public health improvements in health advances: The twentieth-century United States. Demography, 42.
- Davis, D. R. and Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. American Economic Review, 92(5):1269–1289.
- Davis, G. W. (1900). Report on Civil Affairs of Porto Rico, 1899. Washington, DC: Government Printing Office.

- Davis, K. (1953). Puerto Rico: A crowded island. Annals of the American Academy of Political and Social Science, 285(1):116–122.
- de Chaisemartin, C. and D’Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9):2964–2996.
- de Jesús Toro, R. (1982). Historia Económica de Puerto Rico. Cincinnati, OH: Southwestern Publishing Co.
- de Tejada Morales, L. (1877). Memoria sobre el establecimiento de una red de ferrocarriles en la isla de puerto rico. Revista de Obras Públicas, 25(16).
- Deaton, A. (2007). Height, health, and development. Proceedings of the National Academy of Sciences of the United States of America, 104(33):13232–13237.
- del Río, J. (1938). Midwife’s supervision in the rural areas of Puerto Rico. Puerto Rico Health Bulletin, 2(4):171–174.
- Descartes, S. (1946). Basic Statistics on Puerto Rico. Washington, DC: Office of Puerto Rico.
- D’Estefano, M. A. (1968). Puerto Rico: Analysis of a Plebiscite. Havana, Cuba: Tricontinental.
- Devereux, J. (2019). Arrested development? Puerto Rico in an American century. Journal of Economic History, 79(3):708–735.
- Dheda, K. and Migliori, G. B. (2012). The global rise of extensively drug-resistant tuberculosis: is the time to bring back sanatoria now overdue? The Lancet, 379(9817):773–775.
- Dietz, J. L. (1986). Economic History of Puerto Rico: Institutional Change and Capitalist Development. Princeton, NJ: Princeton University Press.
- Dietz, J. L. (2003). Puerto Rico: Negotiating Development and Change. Boulder, CO: Lynne Rienner Publishers, Inc.
- Diffie, B. W. (1931). Porto Rico: A Broken Pledge. New York: The Vanguard Press.
- Domenech, M. V. (1919). Roads in Porto Rico. Overand Monthly and Out West Magazine, 73(4).
- Domínguez, J. I. (1978). Cuba: Order and Revolution. Cambridge, MA: Belknap Press.
- Donaldson, D. and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. Journal of Economic Perspectives, 30(4):171–198.
- Duchesne, R. (2000). The industrialization of Puerto Rico, 1948-1980: Between classical and dependency theory. Journal of Developing Societies, 16(2):223–244.
- Dyer, D. R. (1957). Urbanism in Cuba. Geographical Review, 47(2):292–314.
- Eliot, M. M. (1933). The Effect of Tropical Sunlight on the Development of Bones of Children in Puerto Rico: A Roentgenographic and Clinical Study of Infants and Young Children with Special Reference to Rickets and Related Factors. Washington, DC: Government Printing Office.
- Eriksson, K., Niemesh, G. T., and Thomasson, M. (2018). Revising infant mortality rates for the early twentieth century United States. Demography, 55(6):2001–2024.

- Espinosa, M. (2000). *Sanitary and American: Disease Eradication Efforts and the Transformation of Puerto Rico after 1898*. Latin American Studies Association International Congress, Miami, FL.
- Fageda, X. and Olivieri, C. (2019). Transport infrastructure and regional convergence: A spatial panel data approach. *Papers in Regional Science*, 98(4):1609–1631.
- Feliciano, Z. M. and Green, A. (2017). *US Multinationals in Puerto Rico and the Repeal of Section 936 Tax Exemption for US Corporations*. (NBER Working Paper 23681). Cambridge, MA.
- Fernández, L. (2010). Of immigrants and migrants: Mexican and Puerto Rican labor migration in comparative perspective, 1942-1964. *Journal of American Ethnic History*, 29(3):6–39.
- Fernández, N. A., Burgos, J. C., Ansenjo, C. F., and Rosa, I. (1971). Nutritional status of the Puerto Rican population: Master sample survey. *American Journal of Clinical Nutrition*, 24(8):952–965.
- Fernihough, A. and McGovern, M. E. (2015). Physical stature decline and the health status of the elderly population in England. *Economics and Human Biology*, 16:30–44.
- Ferrie, J. P. and Troesken, W. (2008). Water and Chicago's mortality transition, 1850-1925. *Explorations in Economic History*, 45(1):1–16.
- Fleagle, F. K. (1917). *Results of the Anthropometric Measurements Among the Students at the University of Porto Rico*. University of Porto Rico.
- Fogel, R. W. (1994). Economic growth, population theory, and physiology: The bearing of long-term processes on the making of economic policy. *American Economic Review*, 84(3):369–395.
- Fogel, R. W. (2004). *The Escape from Hunger and Premature Death, 1700–2100: Europe, America, and the Third World*. Cambridge University Press.
- Garcia-Palmieri, M. R., Costas, R., Cruz-Vidal, M., Sorlie, P. D., Tillotson, J., and Havlik, R. J. (1984). Milk consumption, calcium intake, and decreased hypertension in Puerto Rico: Puerto Rico Heart Health Program. *Hypertension*, 6(3):322–328.
- García-Palmieri, M. R., Feliberti, M., Costas Jr., R., Colón, A. A., Cruz-Vidal, M., Cortés-Alicea, M., Ayala, A. M., Sobrino, R., and Torres, R. (2002). An epidemiological study on coronary heart disease in PR: The Puerto Rico Heart Health Program. *Boletín de la Asociación Médica de Puerto Rico*, 94:61–67.
- Garrido Morales, E. (1935). Program for tuberculosis control in Puerto Rico. *JAMA*, 105(23):1842–1844.
- Garrido Morales, E. (1937). Milk stations. *Puerto Rico Health Bulletin*, 1(9):12–15.
- Garver, E. S. and Fincher, E. B. (1945). *Puerto Rico: Unsolved Problem*. Elgin, IL: Brethren Publishing House.
- Godoy, R. A., Goodman, E., Levins, R., Caram, M., and Seyfried, C. (2007). Adult male height in an American colony: Puerto Rico and the USA mainland compared, 1886-1955. *Economics and Human Biology*, 5(1):82–99.

- Goldenberg, R., McClure, E., Saleem, S., and Reddy, U. (2010). Infection-related stillbirths. The Lancet, 375(9724):1482–1490.
- González, E. M. (2015). Nurturing the citizens of the future: Milk stations and child nutrition in Puerto Rico, 1929-60. Medical History, 59(2):177–198.
- Goodman-Bacon, A. (forthcoming). Difference-in-differences with variation in treatment timing. Journal of Econometrics.
- Graff, T. (2019). Spatial Inefficiencies in Africa’s Trade Network. (NBER Working Paper No. 25951). Cambridge, MA: National Bureau of Economic Research.
- Grau, H. R., Aide, T. M., Zimmerman, J. K., and Thomlinson, J. R. (2004). Trends and scenarios of the carbon budget in postagricultural Puerto Rico (1936-2060). Global Change Biology, 10:1163–1179.
- Greely, A. (1900). Report on Military Telegraph Lines in Porto Rico. Washington, DC: Government Printing Office.
- Guimaraes, P., Rolfe, R. J., and Woodward, D. P. (1998). Regional incentives and industrial location in Puerto Rico. International Regional Science Review, 21(2):119–138.
- Hanlon, W. W. (2017). Temporary shocks and persistent effects in urban economies: Evidence from British cities after the US Civil War. Review of Economics and Statistics, 99(1):67–79.
- Hansen, C. W., Jensen, P. S., and Madsen, P. E. (2020). Preventing the white death: Tuberculosis dispensaries. Economic Journal, 130(629):1288–1316.
- Hardoy, J. E. (1975). Two thousands years of Latin American urbanization. In J.E. Hardoy (Ed.), *Urbanization in Latin America: Approaches and Issues* (pp. 3-56). Garden City, NY: Doubleday & Company, Inc.
- Hardy, A. and Kramer, M. (1941). The reporting of mortality in Puerto Rico. Puerto Rico Health Bulletin, 5(1):6–10.
- Harris, B. (2007). Growing taller, living longer? Anthropometric history and the future of old age. Ageing and Society, 17(5):491–512.
- Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. Annals of the Association of American Geographers, 44(4):315–348.
- Hatton, T. J. (2014). How have Europeans grown so tall? Oxford Economic Papers, 66:349–372.
- Hatton, T. J. and Bray, B. E. (2010). Long run trends in the heights of European men, 19th-20th centuries. Economics and Human Biology, 8:405–413.
- Henderson, J. V. (2010). Cities and development. Journal of Regional Science, 50(1):515–540.
- Hoehn-Velasco, L. (2018). Explaining declines in US rural mortality, 1910-1933: The role of county health departments. Explorations in Economic History, 70:42–72.
- Hoehn-Velasco, L. and Wrigley-Field, E. (2021). Do Public Health Departments Improve Population Health? The Impact of City-level Health Departments over 1916-1933. (Working Paper).

- Hollingsworth, A. (2014). Controlling TB in a World without Antibiotics: Isolation and Education in North Carolina, 1932-1940. (Working Paper).
- Horan, P. M. and Hargis, P. G. (1995). County Longitudinal Template, 1840-1990: Version 1 [dataset]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. Data retrieved from <https://doi.org/10.3886/ICPSR06576.v1>.
- Hotez, P. J., Brooker, S., Bethony, J. M., Bottazzi, M. E., Loukas, A., and Xiao, S. (2004). Current concepts: Hookworm infection. New England Journal of Medicine, 351(8):799–807.
- Huillery, E. (2009). History matters: The long-term impact of colonial public investments in French West Africa. American Economic Journal: Applied Economics, 1(2):176–215.
- Hull, H. (1915). Porto Rico's modern highways. Motor Age, 27(7):33–35.
- Immerwahr, D. (2019). How to Hide an Empire: A History of the Greater United States. New York: Farrar, Straus and Giroux.
- International Urban Research (1959). The World's Metropolitan Areas. Berkeley, CA: University of California Press.
- Jackson, J. F. (2020). Roads to American empire: US military public works and capitalist transitions, 1898-1934. Journal of Historical Sociology, 33(1):116–133.
- Jackson, P. (1980). Social and spatial aspects of Puerto Rican migration. Singapore Journal of Tropical Geography, 1(1):37–45.
- Jana, N., Vasishta, K., Jindal, S., Khunnu, B., and Ghosh, K. (1994). Perinatal outcome in pregnancies complicated by pulmonary tuberculosis. International Journal of Gynecology and Obstetrics, 44(2):119–124.
- Janer, J. L. (1945). Population growth in Puerto Rico and its relation to time changes in vital statistics. Human Biology, 17(4):267–313.
- Janer, J. L. (1958). Vital Statistics System of the Commonwealth of Puerto Rico. (Mimeo).
- Jaworski, T. and Kitchens, C. T. (2019). National policy for regional development: Evidence from Appalachian highways. Review of Economics and Statistics, 101(5):777–790.
- Jayachandran, S., Lleras-Muney, A., and Smith, K. V. (2010). Modern medicine and the twentieth century decline in mortality: Evidence on the impact of sulfa drugs. American Economic Journal: Applied Economics, 2(2):118–146.
- Jedwab, R., Kerby, E., and Moradi, A. (2015). History, path dependence and development: Evidence from colonial railways, settlers and cities in Kenya. Economic Journal, 127(603):1467–1494.
- Jedwab, R. and Moradi, A. (2016). The permanent effects of transportation revolutions in poor countries: Evidence from Africa. Review of Economics and Statistics, 98(2):268–284.
- Jedwab, R. and Storeygard, A. (2018). The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa. (Working Paper).

- Kelly, M. (2019). The Standard Errors of Persistence. (Working Paper).
- Kitchens, C. (2013). The effects of the Works Progress Administration's anti-malaria programs in Georgia, 1932-1947. Explorations in Economic History, 50(4):567–581.
- Klein, A. and Crafts, N. (2012). Making sense of the manufacturing belt: Determinants of US industrial location, 1880-1920. Journal of Economic Geography, 12(4):775–807.
- Knott, V. B. (1963). Stature, leg girth, and body weight of Puerto Rican private school children. Growth, 27:157–174.
- Komisarow, S. (2017). Public health regulation and mortality: Evidence from early 20th century milk laws. Journal of Health Economics, 56:126–144.
- Komlos, J. (2001). On the biological standard of living of eighteenth-century Americans: Taller, richer, healthier. Research in Economic History, 20:223–248.
- Komlos, J. and Baten, J., editors (1998). The Biological Standard of Living in Comparative Perspective. Stuttgart, Germany: Franz Steiner Verlag.
- Kotsadam, A., Lind, J. T., and Modalsi, J. (2017). Call the Midwife: Health Personnel and Mortality in Norway 1887-1921. (CESifo Working Paper Series No. 6831).
- Krugman, P. (1991). Increasing returns and economic geography. Journal of Political Economy, 99(3):483–9.
- Laska-Mierzejewska, T. (1970). Morphological and developmental difference between Negro and white Cuban youths. Human Biology, 42(4):581–597.
- Lazuka, V. (2018). The long-term health benefits of receiving treatment from qualified midwives at birth. Journal of Development Economics, 133:415–433.
- Lee, R. (2003). The demographic transition: Three centuries of fundamental change. Journal of Economic Perspectives, 17(4):167–190.
- Lefort, F. (1997). Is Puerto Rico converging to the United States? (International Tax Program, Harvard Law School, Report No. 1003). Cambridge, MA.
- Lewis, G. K. (1963). Puerto Rico: Freedom and Power in the Caribbean. New York: MR Press.
- Lindhardt, M. (1939). The Statistics of Pulmonary Tuberculosis in Denmark, 1925-34. Copenhagen: Munksgaard.
- López-Alonso, M. and Porrás Condey, R. (2003). The ups and downs of Mexican economic growth: The biological standard of living and inequality, 1870-1950. Economics and Human Biology, 1(2):169–186.
- Mahapatra, P., Shibuya, K., Lopez, A. D., Coullare, F., Notzon, F. C., Rao, C., and Szreter, S. (2007). Civil registration systems and vital statistics: Successes and missed opportunities. The Lancet, 370(9599):1653–1563.
- Malaret, P. (1937). Malaria. Puerto Rico Health Bulletin, 1(7):148–152.

- Martín Ramos, J. (2005). Las Comunicaciones en la Isla de Puerto Rico: 1850-1898. San Juan, PR: Academia Puertorriqueña de la Historia.
- Martinez Carrión, J. M. and Pérez Castejón, J. J. (1998). Height and standards of living during the industrialization of Spain: The case of Elche. European Review of Economic History, 2(2):201–230.
- Martinez-Galarraga, J., Paluzie, E., Pons, J., Silvestre, J., and Tirado, D. (2020). New economic geography and economic history: A survey of recent contributions through the lens of the Spanish industrialization process. Cliometrica.
- Martínez Rivera, E. (1943). A note on the development of public health in Puerto Rico. Boletín de la Asociación Médica de Puerto Rico, 35(10):384–391.
- Martinuzzi, S., Gould, W. A., and Ramos González, O. M. (2007). Land development, land use, and urban sprawl in Puerto Rico integrating remote sensing and population census data. Landscape and Urban Planning, 79:288–297.
- McEniry, M. (2011). Infant mortality, season of birth and the health of older Puerto Rican adults. Social Science and Medicine, 72(6):1004–1015.
- McEniry, M. (2014). Early Life Conditions and Rapid Demographic Changes in the Developing World. Dordrecht, Netherlands: Springer.
- McKeown, T. (1979). The Role of Medicine: Dream, Mirage, or Nemesis? Oxford: Basil Blackwell.
- Meisel, A. and Vega, M. (2007). The biological standard of living (and its convergence) in Colombia, 1870-2003: A tropical success story. Economics and Human Biology, 5(1):100–122.
- Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. Review of Economics and Statistics, 90(4):683–701.
- Michaels, G., Rauch, F., and Redding, S. J. (2012). Urbanization and structural transformation. Quarterly Journal of Economics, 127(2):535–586.
- Miller, P. G. (1922). Historia de Puerto Rico. Chicago, IL: Rand McNally & Company.
- Mintz, S. W. (1974). Caribbean Transformations. Chicago, IL: Aldine Publishing Company.
- Miranda Franco, R. and Casta Vélez, A. (1997). La erradicación de la malaria en Puerto Rico. Revista Panamericana de Salud Pública, 2(2):146–150.
- Mitchener, K. J. and Weidenmier, M. (2005). Empire, public goods, and the Roosevelt Corollary. Journal of Economic History, 65(3):658–692.
- Moehling, C. M. and Thomasson, M. A. (2014). Saving babies: The impact of public education programs on infant mortality. Demography, 51(2):367–386.
- Molina, G. and Noam, I. F. (1964). Indicators of health, economy and culture in Puerto Rico and Latin America. American Journal of Public Health and the Nation's Health, 54(8):1191–1206.
- Moore, K. A., Simpson, J. A., Scoullar, M. J., McGready, R., and Rowkes, F. J. (2017). Quantification of the association between malaria in pregnancy and stillbirth: A systematic review and meta-analysis. The Lancet Global Health, 5(11):PE1101–E1112.

- Moradi, A. (2009). Towards an objective account of nutrition and health in colonial Kenya: A study of stature in African army recruits and civilians, 1880-1980. Journal of Economic History, 69(3):719–754.
- Morales Carrión, A. (1983). Puerto Rico: A Political and Cultural History. New York: W.W. Norton and Company, Inc.
- Morgan, S. L. and Liu, S. (2001). Was Japanese colonialism good for the welfare of Taiwanese? Stature and the standard of living. The China Quarterly, 192:990–1017.
- Morning Consult (2017). National tracking poll #170916. September 22-24, [Survey report] Retrieved from [https://morningconsult.com/wp-content/uploads/2017/10/170916\\_crosstabs\\_pr\\_v1\\_KD.pdf](https://morningconsult.com/wp-content/uploads/2017/10/170916_crosstabs_pr_v1_KD.pdf).
- Mountin, J. W., Pennell, E. H., and Flook, E. (1937). Illness and Medical Care in Puerto Rico. Washington, DC: Government Printing Office.
- Nagy, D. (2018). Trade and Urbanization: Evidence from Hungary. (Working Paper).
- NCD Risk Factor Collaboration (NCD-RisC) (2016). A century of trends in adult human height. eLife, 5.
- Nochera, D. (1940). The public health unit as an educational entity. Puerto Rico Health Bulletin, 4(3):112–116.
- Núñez, J. (2005). Signed with an X: Methodology and data sources for analyzing the evolution of literacy in Latin America and the Caribbean, 1900-1950. Latin American Research Review, 40(2):117–135.
- Okoye, D., Pongou, R., and Yokossi, T. (2019). New technology, better economy? the heterogeneous impact of colonial railroads in nigeria. Journal of Development Economics, 140:320–354.
- Olds, K. B. (2001). The biological standard of living in Taiwan under Japanese occupation. Economics and Human Biology, 1(2):187–206.
- Olmstead, A. L. and Rhode, P. W. (2004). An impossible undertaking: The eradication of bovine tuberculosis in the United States. Journal of Economic History, 64(3):734–772.
- Ortiz Colom, J. (2015). La carretera de Cayey a Arroyo por Guayama. In *Actas del Noveno Congreso Nacional y Primer Congreso Internacional Hispanoamericano de Historia de Construcción* (pp. 1259-1268). Madrid: Instituto Juan de Herrera, Universidad Politécnica de Madrid.
- Palloni, A., Davila, A. L., and Sanchez-Ayendez, M. (2013). Puerto Rican Elderly: Health Conditions (PREHCO) Project, 2002-2003, 2006-2007 [dataset]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Data retrieved from <https://www.icpsr.umich.edu/icpsrweb/DSDR/studies/34596>.
- Palsson, C. (2020). “A Whirligig of Revolutionary Presidents”: Political Stability and Foreign Investment in Haiti, 1910-1920. (Working Paper).
- Payne, G. C. (1942). Some considerations concerning the development of public health units in Puerto Rico. Puerto Rico Health Bulletin, 6(9):265–270.

- Perkins, J. M., Subramanian, S., Smith, G. D., and Özaltin, E. (2016). Adult height, nutrition, and population health. *Nutrition Reviews*, 74(3):149–165.
- Perloff, H. S. (1950). *Puerto Rico's Economic Future: A Study in Planned Development*. Chicago, IL: University of Chicago Press.
- Perloff, H. S. (1952). The United States and the economic development of Puerto Rico. *Journal of Economic History*, 12:45–59.
- Peter, R. S., Fromm, E., Klenk, J., Concin, H., and Nagel, G. (2014). Change in height, weight, and body mass index: Longitudinal data from Austria. *American Journal of Human Biology*, 26(5):690–696.
- Pettersson-Lidbom, P. (2015). *Midwives and Maternal Mortality: Evidence from a Midwifery Policy Experiment in Sweden in the 19th Century*. (Working Paper.)
- Picó, F. (1983). El impacto de la invasión americana en la zona cafetalera de Puerto Rico: El caso de Utuado. In B.G. Silvestrini (Ed.), *Politics, Society, and Culture in the Caribbean: Selected Papers of the XIV Conference of Caribbean Historians* (pp. 133-144). San Juan, PR: Universidad de Puer—to Rico.
- Picó, R. (1969). *Nueva Geografía de Puerto Rico*. San Juan, PR: Editorial Universitaria, Universidad de Puerto Rico.
- Picó, R. (1974). *The Geography of Puerto Rico*. Chicago: Aldine Publishing Company.
- Pons, J. A. (1952). Las unidades de salud publica de Puerto Rico. *Boletin de la Oficina Sanitaria Panamericana*, 33(1):1–10.
- Prados de la Escosura, L. (2015). Human development as positive freedom: Latin America in historical perspective. *Journal of Human Development and Capabilities*, 16:342–373.
- Preston, S. H. (1980). Causes and consequences of mortality declines in less developed countries during the twentieth century. In R.A. Easterlin (Ed.), *Population and Economic Change in Developing Countries* (pp. 289-360). Chicago, IL: University of Chicago Press.
- Puerto Rico. Commissioner of Health (1926). *Report of the Commissioner of Health of Porto Rico to the Governor of Porto Rico for the Fiscal Year Ending June 30, 1926*. San Juan, PR: Bureau of Supplies, Printing, and Transportation.
- Puerto Rico. Commissioner of Health (1932). *Report of the Commissioner of Health of Puerto Rico to the Hon. Governor of Puerto Rico for the Fiscal Year 1931-32*. San Juan, PR: Bureau of Supplies, Printing, and Transportation.
- Puerto Rico. Commissioner of Health (1934). *Report of the Commissioner of Health of Puerto Rico to the Hon. Governor of Puerto Rico for the Fiscal Year 1933-34*. San Juan, PR: Bureau of Supplies, Printing, and Transportation.
- Puerto Rico. Commissioner of Health (1935). *Report of the Commissioner of Health of Puerto Rico to the Governor of Puerto Rico for the Fiscal Year Ending June 30, 1935*. San Juan, PR: Bureau of Supplies, Printing, and Transportation.

- Puerto Rico. Commissioner of Health (1936). Report of the Commissioner of Health of Puerto Rico to the Hon. Governor of Puerto Rico for the Fiscal Year 1935-36. San Juan, PR: Bureau of Supplies, Printing, and Transportation.
- Puerto Rico. Commissioner of Health (1938). Report of the Commissioner of Health of Puerto Rico to the Hon. Governor of Puerto Rico for the Fiscal Year 1937-38. San Juan, PR: Bureau of Supplies, Printing, and Transportation.
- Puerto Rico. Commissioner of Health (1939). Report of the Commissioner of Health of Puerto Rico to the Hon. Governor of Puerto Rico for the Fiscal Year 1938-39. San Juan, PR: Bureau of Supplies, Printing, and Transportation.
- Puerto Rico. Commissioner of Health (1942). Report of the Commissioner of Health to the Hon. Governor of Puerto Rico for the Fiscal Year 1941-42. San Juan, PR: Bureau of Supplies, Printing, and Transportation.
- Puerto Rico. Department of Labor and Human Resources (2010). Serie Histórica de Empleo y Desempleo Puerto Rico. Retrieved from <https://estadisticas.pr>.
- Puerto Rico Endowment for the Humanities and the National Endowment for the Humanities (2018). Municipalities. In *Encyclopedia of Puerto Rico*. Retrieved from <https://enciclopediapr.org/en/home/municipalities/>.
- Puerto Rico. Governor (1901). First Annual Report of Charles H. Allen, Governor of Porto Rico, Covering the Period from May 1, 1900 to May 1, 1901. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1902). Second Annual Report of the Governor of Porto Rico, Covering the Period from May 1, 1901 to July 1, 1902. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1904). Fourth Annual Report of the Governor of Porto Rico. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1907). Annual Report of the Governor of Porto Rico for the Fiscal Year Ending June 30, 1907. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1910). Report of the Governor of Porto Rico to the Secretary of War. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1919). Report of the Governor of Porto Rico to the Secretary of War. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1923). Twenty-third Annual Report of the Governor of Porto Rico. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1924). Twenty-fourth Annual Report of the Governor of Porto Rico. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1926). Twenty-fifth Annual Report of the Governor of Porto Rico. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1929). Twenty-ninth Annual Report of the Governor of Porto Rico. Washington, DC: Government Printing Office.

- Puerto Rico. Governor (1932). Thirty-First Annual Report of the Governor of Porto Rico. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1935). Thirty-fifth Annual Report of the Governor of Puerto Rico. Washington, DC: Government Printing Office.
- Puerto Rico. Governor (1938). Thirty-Eighth Annual Report of the Governor of Puerto Rico. San Juan, PR: Bureau of Supplies, Printing, and Transportation.
- Puerto Rico. Governor (1942). Forty-Second Annual Report of the Governor of Puerto Rico. San Juan, PR: Bureau of Supplies, Printing, and Transportation.
- Pumarada O'Neill, L. F. (1995). Land transportation in Puerto Rico, c. 1508-1950. (National Register of Historic Places Multiple Property Documentation Form).
- Quintero, E. (1940). Venereal disease clinics. Puerto Rico Health Bulletin, 34(2):43–51.
- Ramírez de Arellano, A. B. (1981). The politics of public health in Puerto Rico: 1926-40. Revista de Salud Pública de Puerto Rico, 3:35–58.
- Ramos, F. (1992). Out-migration and return migration of Puerto Ricans. In G.J. Borjas and R.B. Freeman (Eds.), *Immigration and the Workforce: Economic Consequences for the United States and Source Areas* (pp. 49-66). Chicago, IL: University of Chicago Press.
- Redding, S. J., Sturm, D. M., and Wolf, N. (2011). History and industry location: Evidence from German airports. Review of Economics and Statistics, 93(3):814–831.
- Rivera-Batiz, F. L. and Santiago, C. E. (1996). Island Paradox: Puerto Rico in the 1990s. New York: Russell Sage Foundation.
- Robert de Romeu, M. (1937). Child and maternal hygiene. Puerto Rico Health Bulletin, 1(5):92–95.
- Robert de Romeu, M. (1938). What we have done to decrease our deaths from tetanus of the newborn. Puerto Rico Health Bulletin, 2(1):55–58.
- Robert de Romeu, M. (1939). Relation of the assistant midwife to public health in Puerto Rico. Puerto Rico Health Bulletin, 3(6):229–235.
- Roberts, R. C. (1941). Soil Survey of Puerto Rico. Washington, DC: Department of Agriculture.
- Robinson, A. G. (1899). The Porto Rico of To-day; Pen Pictures of the People and the Country. New York: Scribner's Sons.
- Rodríguez Pastor, J. and Janer, J. L. (1953). Tuberculosis in the island of Puerto Rico. American Review of Tuberculosis, 67(2):132–153.
- Roessler, P., Pengl, Y., Marty, R., Titlow, K., and van de Walle, N. (2020). Cash Crops, Colonialism and Legacies of Spatial Inequality: Evidence from Africa. (Working Paper).
- Roser, M., Ortiz-Ospina, E., and Ritchie, H. (2020). *Life Expectancy* [Data file]. Retrieved from *Our World in Data*, <https://ourworldindata.org/life-expectancy>.
- Rudel, T. K., Perez-Lugo, M., and Zichal, H. (2000). When fields revert to forest: Development and spontaneous reforestation in post-war Puerto Rico. The Professional Geographer, 52:386–397.

- Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M. (2018). IPUMS USA: Version 8.0 [dataset]. Minneapolis, MN: IPUMS. Data retrieved from <https://doi.org/10.18128/D010.V8.0>.
- Russell, M. (1976). Parent-child and sibling-sibling correlations of height and weight in a rural Guatemalan population of preschool children. Human Biology, 48(3):501–515.
- Sachs, J. D. (2001). Tropical Underdevelopment. (NBER Working Paper 8119). Cambridge, MA: National Bureau of Economic Research.
- Salvatore, R. D. (2004). Stature decline and recovery in a food-rich export economy: Argentina 1900-1934. Explorations in Economic History, 42(1):233–255.
- Samaras, T. T. (2012). How height is related to our health and longevity: A review. Nutrition and Health, 21(4):247–261.
- Sandis, E. E. (1970). Characteristics of Puerto Rican migrants to, and from, the United States. International Migration Review, 4(2):22–43.
- Santamaría García, A. (2012). Economía, 1800-1898. In L.E. González and M. Dolores Luque (Eds.), *Historia de Puerto Rico*, Vol. 4 of *Historia de las Antillas* (pp. 135-167). Aranjuez, Spain: Consejo Superior de Investigaciones Científicas/Ediciones Doce Calles.
- Scarano, F. A. and Curtis, K. J. (2011). Puerto Rico's *padrones*, 1779-1802. data retrieved from Inter-university Consortium for Political and Social Research, <https://doi.org/10.3886/ICPSR30262.v1>.
- Schmidheiny, K. and Siegloch, S. (2020). On Event Study Designs and Distributed-Lag Models: Equivalence, Generalization and Practical Implications. (IZA DP No. 12079). Bonn, Germany: Institute of Labor Economics.
- Silventoinen, K. (2003). Determinants of variation in adult body height. Journal of Biosocial Science, 35(2):263–285.
- Smith, D. (1937). Some Standards for Measuring Puerto Rico's Economic and Social Progress. Washington, DC: Association of Sugar Producers of Puerto Rico.
- Solá, J. O. (2011). Colonialism, planters, sugarcane, and the agrarian economy of Caguas, Puerto Rico, between the 1890s and 1930. Agricultural History, 85(3):349–372.
- Spijker, J. J., Cámara, A. D., and Blanes, A. (2012). The health transition and biological living standards: Adult height and mortality in 20th-century Spain. Economics and Human Biology, 10(3):276–288.
- Steckel, R. H. (1995). Stature and the standard of living. Journal of Economic Literature, 33(4):1903–1940.
- Steckel, R. H. (2009). Heights and human welfare: Recent developments and new directions. Explorations in Economic History, 46(1):1–23.
- Stolnitz, G. J. (1953). Recent international differences and trends in expectation of life. Population Index, 19(1):2–10.

- Stone, L. O. (1965). Population redistribution and economic development in Puerto Rico, 1950-1960. *Social and Economic Studies*, 14(3):264–271.
- Straub, S. (2015). Policy lessons from the recent literature on transport infrastructure development. In A. Picot, M. Florio, N. Grove, and J. Kranz (Eds.), *The Economics of Infrastructure Provisioning: The Changing Role of the State* (pp. 267-284). Cambridge, MA: MIT Press.
- Tavares Da Silva, F., Gonik, B., McMillan, M., Keech, C., Dellicour, S., Bhange, S., Tila, M., Harper, D. M., Woods, C., Kawai, A. T., Kochhar, S., and Munoz, F. M. (2016). Stillbirth: Case definition and guidelines for data collection, analysis, and presentation of maternal immunization safety data. *Vaccine*, 34(49):6057–6068.
- Taylor, M. C. (1952). Neo-Malthusiansm in Puerto Rico. *Review of Social Economy*, 10(1):42–54.
- Thieme, F. P. (1957). A comparison of Puerto Rican migrants and *sedentes*. *Papers of the Michigan Academy of Science, Arts, and Letters*, 42:249–256.
- Thieme, F. P. (1959). *The Puerto Rican Population: A Study in Human Biology*. (Museum of Anthropology, University of Michigan, Anthropological Papers No. 13). Ann Arbor, MI.
- Thomasson, M. A. and Treber, J. (2008). From home to hospital: The evolution of childbirth in the United States, 1928-1940. *Explorations in Economic History*, 45(1):76–99.
- Trujillo-Pagan, N. (2013). *Modern Colonization by Medical Intervention: US Medicine in Puerto Rico*. Leiden, Netherlands: Brill.
- United States. Bureau of Labor Statistics (1943). Wages and hours in the needlework industry in Puerto Rico, 1943. *Monthly Labor Review*, 61(4):778–779.
- United States. Bureau of Labor Statistics (2020). *Historical Consumer Price Index for All Urban Consumers (CPI-U): US City Average, All Items* [Data file]. Retrieved from <https://www.bls.gov/cpi>.
- United States. Department of Commerce and Labor (1907). *Commercial Porto Rico in 1906. Showing commerce, production, transportation, finances, area, population, and details of trade with the United States and foreign countries during a term of years*. Washington, DC: Government Printing Office.
- United States. Department of Health and Human Services (1985). Hispanic Health and Nutrition Examination Survey, 1982-1984 (HHANES) [dataset]. Hyattsville, MD: Centers for Disease Control and Prevention. Data retrieved from <https://wwwn.cdc.gov/nchs/nhanes/Hhanes/Default.aspx>.
- United States. Interdepartmental Committee on Nutrition for National Defense (1964). *Venezuela: Nutrition Survey, May-June 1963; A Report*. Washington, DC.
- United States. National Center for Chronic Disease Control (1967). *Republic of Paraguay; Nutrition Survey of the Armed Forces, May-August 1965*. Bethesda, MD.
- United States. Tariff Commission (1943). *Puerto Rico's Economy with Special Reference to United States-Puerto Rican Trade* [Mimeo]. San Juan, PR: Bureau of Supplies, Printing, and Transportation.

- United States. War Department, Adjutant-General's Office, Military Information Division (1898). Military Notes on Puerto Rico. Washington, DC: Government Printing Office.
- Vaughan, S. and Schwirian, K. (1979). A longitudinal study of metropolitan density patterns in a developing country: Puerto Rico, 1899-1970. Environment and Planning A: Economy and Space, 11(4):423-433.
- Vázquez Calzada, J. L. (1968). Fertility decline in Puerto Rico: Extent and causes. Demography, 5(2):855-865.
- Vázquez Calzada, J. L. (1978). La Población de Puerto Rico y su Trayectoria Histórica. Río Piedras, PR: Escuela Graduada de Salud Pública, Recinto de Ciencias Médicas, Universidad de Puerto Rico.
- Vázquez Calzada, J. L. (1984). El descenso de la mortalidad en Puerto Rico. Puerto Rico Health Sciences Journal, 3(4):173-182.
- Vilar Isern, R. A. (1935). Memorial de la Asociacion Medica de Puerto Rico en relacion con la benefica municipal. Boletín de la Asociación Médica de Puerto Rico, 27(3):71-79.
- Ward, M. and Devereux, J. (2012). The road not taken: Pre-revolutionary Cuban living standards in comparative perspective. Journal of Economic History, 72(1):104-133.
- Washburn, B. (1939). The hookworm campaign (a story). Puerto Rico Health Bulletin, 3(9):389-393.
- Wegman, M. E., Fernández Marchante, R., and Kramer, M. (1942). Infant mortality and infant feeding in Puerto Rico. Puerto Rico Journal of Public Health and Tropical Medicine, 17:228-245.
- Willoughby, W. F. (1905). Territories and Dependencies of the United States: Their Government and Administration. New York: The Century Co.
- Wilson, S. E. (2011). Chasing success: Health sector aid and mortality. World Development, 39(11):2032-2043.
- World Bank (2011). World Development Indicators 2011. Washington, DC: World Bank.
- World Economic Forum (2015). The Travel and Tourism Competitiveness Report 2015. Geneva: World Economic Forum.
- World Health Organization (2000). Effect of breastfeeding on infant and child mortality due to infectious diseases in less developed countries: A pooled analysis. The Lancet, 355(9202):451-455.
- Wüst, M. (2012). Early interventions and infant health: Evidence from the Danish home visiting program. Labour Economics, 19(4):484-495.
- Yelton, S. E. (1946). Tuberculosis throughout the world: 1. the prewar distribution of tuberculosis throughout the world. Public Health Reports, 61(31):1144-1160.
- Zimran, A. (2020). Transportation and health in the antebellum United States, 1820-1847. Journal of Economic History, 80(3):670-709.

## Appendix A

### Supplemental Material for Chapter 2

#### A.1 Robustness Checks

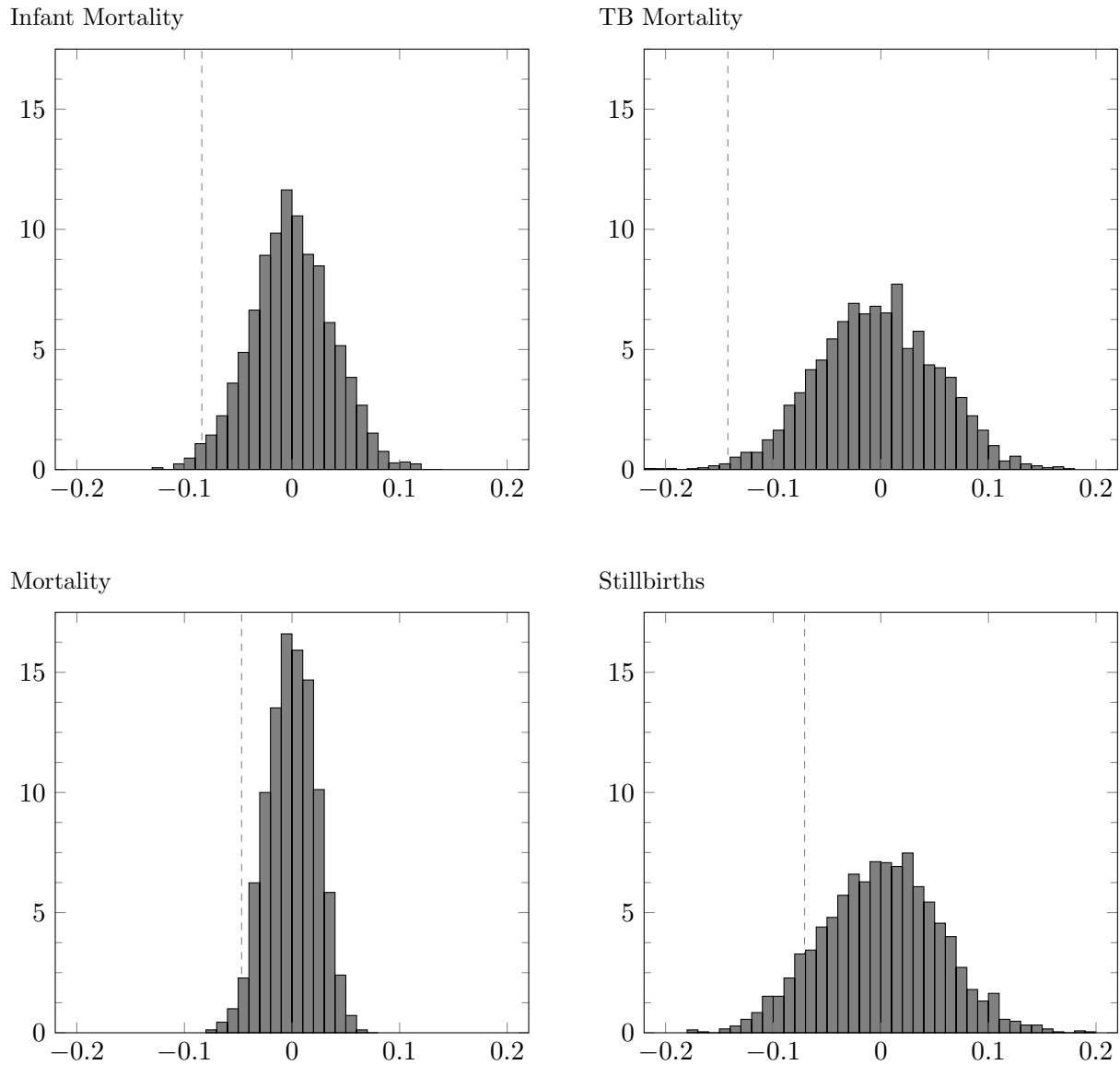
The regression results described in Section 2.5 suggest that PHUs played a major role in the decline in mortality rates in Puerto Rico through the end of World War II. In this subsection, I explore the robustness of the results to several tweaks to the data and econometric model.

##### A.1.1 Random Placebo Test

The evidence from the event study that PHUs significantly reduced mortality applies to four different outcome variables, so it is unlikely that the results are spurious. Nonetheless, in this section I consider the possibility that the relationship between PHUs and mortality is spurious by performing random placebo tests. In each of 2,500 simulations, municipalities are randomly assigned an inaugural treatment year between 1927 and 1938, and the placebo treatment is used to re-estimate Equation 2.1. Figure A.1 shows the distribution of coefficients from these simulations, with the coefficient derived from the true rollout marked by the vertical dashed line. Each distribution is centered near zero, and the percent of placebo coefficients to the left of the dashed lines is 1.28% for infant mortality, 0.56% for TB mortality, 2.12% for general mortality, and 10.08% for stillbirths. Hence, the estimated effect of PHUs is much more negative than would be expected from a placebo intervention and the results are unlikely to be spurious for infant, TB, and general mortality. In the case of stillbirths, the true coefficient falls in far left of the distribution, but the test leaves open the possibility that the negative effect found in Section 2.5 is spurious. The

estimated effect of PHUs on stillbirths is not statistically significant in this specification, anyway.

Figure A.1: The Distributions of Coefficients From Random Placebo Tests



Notes: Each histogram shows the distribution of estimated effects from a placebo intervention with a staggered rollout in which all municipalities are randomly assigned an inaugural year between 1927 and 1938. The distribution summarizes the results from 2,500 simulations. The estimated effect of public health units is represented by a dashed vertical line. Effects are estimated from Equation 2.1 and averaging across point estimates for event-years 0 to 7.

### A.1.2 Underregistration of Births

Births were underreported in Puerto Rico throughout the period under investigation. The US Bureau of the Census in 1931 made tests of the completeness of birth registration in Puerto Rico to determine whether the island was eligible for admission to the US birth registration area. Registration was found to be poor, and the tests were discontinued (Eliot, 1933). Tests of completeness of birth registration based on the 1940 and 1950 censuses concluded that 86 and 96 percent of births, respectively, were registered (Vázquez Calzada, 1968). Puerto Rico was not admitted to the US birth registration area until 1943.

Infant mortality rates suffer from an upward bias stemming from underregistration of births, and the official infant mortality rate may have declined over time as a result of improvements in birth registration. If underregistration varied across municipalities—as in the US, where underreporting was more severe in the South and among blacks—estimates may be biased even in a panel setting with specifications that include location fixed effects and place-specific linear time trends (Eriksson et al., 2018).

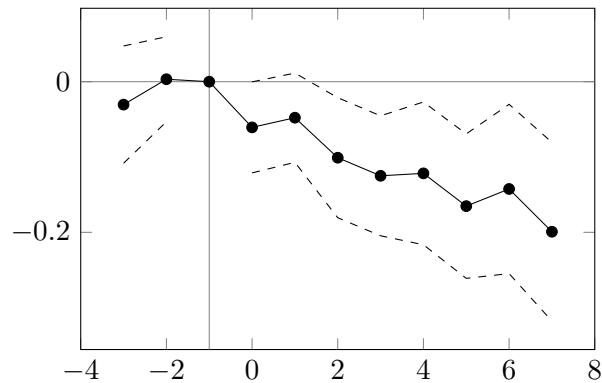
To assess whether the decline in the official infant mortality rate following PHU openings is the product of improvements in birth registration, I here use the ratio of infant deaths to population as the dependent variable. Unlike births, nearly all deaths were believed to be reported even prior to the creation of the Bureau of Vital Statistics in 1931 (Janer, 1945). Likewise, intercensal population counts, though obviously imprecise, are untainted by underregistration of births because they are interpolations of growth trends between censuses. Therefore, the ratio of infant deaths to population has the advantages of being highly correlated with the infant mortality rate ( $\rho = 0.92$ ) but uncorrelated with improvements in birth registration.

The results using this dependent variable appear in Figure A.2 and are nearly identical to results based on the infant mortality rate. Parallel pre-trends demonstrate that changes in infant mortality were not different in other municipalities leading up to program adoption and support the assumption that infant mortality would likewise have followed a parallel trend in the absence

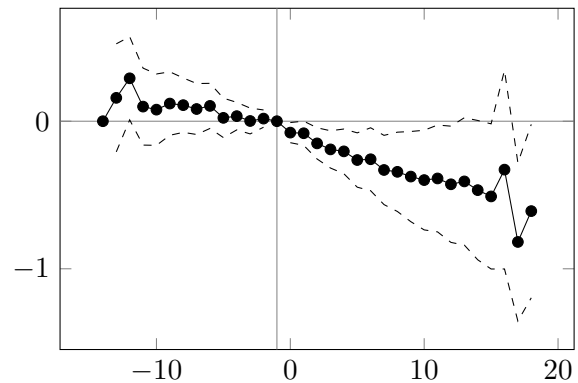
of a PHU. The ratio of infant deaths to population then falls by six or seven percent, depending on the specification, in the first year with a PHU, significant at the 5 percent level. The effect of PHUs then grows over time. Hence, the estimated reduction in infant mortality resulting from PHUs is not driven by differential changes in birth registration.

Figure A.2: The Effect of Public Health Units on the Ratio of Infant Deaths to Population

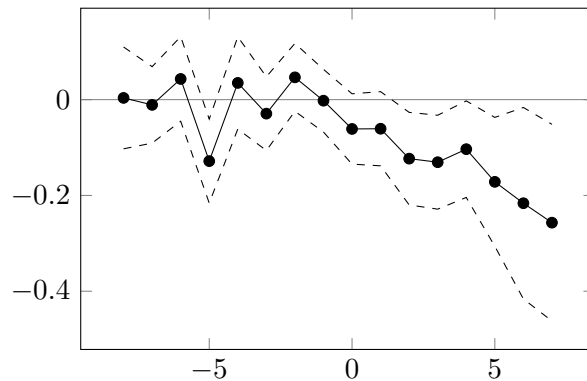
(a) Baseline Specification



(c) Dynamic Specification



(b) de Chaisemartin and D'Haultfoeuille (2020) Estimator



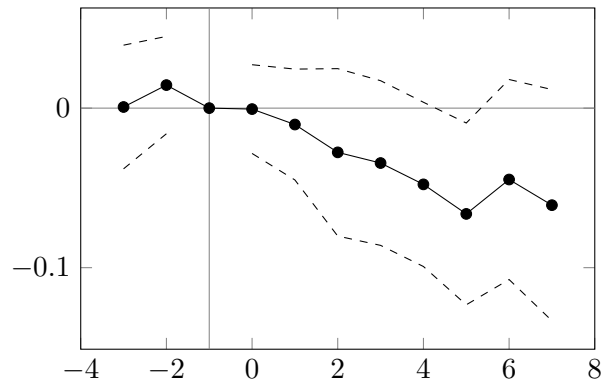
Notes: The dependent variables is the natural log of the ratio of deaths under 1 year old to population, computed by multiplying the infant mortality rate by the birth rate. Coefficients are weighted least-squares estimates. Dashed lines are 95 percent confidence intervals and are based on standard errors clustered by municipality. The year prior to the establishment of the PHU is omitted in Panel A, as is the first event-year in the sample in Panel B. Sample includes 76 municipalities and 1,672 municipality-year observations. In Panel A, event-years outside of the designated window are captured by “before” and “after” dummies.

### A.1.3 Changes in Fertility

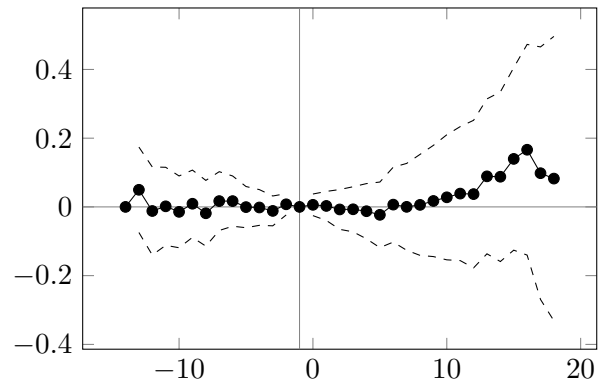
There may be concern that the results presented in the body of the paper attributing reductions in infant mortality to PHUs may be driven by falling fertility, for example by lengthening birth intervals or shifting the composition of births to lower parity. In this section I consider whether the birth rate responds to the rollout of PHUs. Figure A.3 displays results in which the dependent variable is the log fertility rate. There is a downward trend in fertility in the baseline specification, although the estimated effect is only significant at the 10-percent level in the fifth year with a PHU and at the 5-percent level in the sixth year. The dynamic specification and estimator from de Chaisemartin and D'Haultfoeuille (2020), however, offer no indication that fertility changed in response to PHUs. Likewise, unweighted results (not shown) do not suggest that fertility changed. Therefore, the decline in infant mortality in response to PHUs is not the result of changes in fertility.

Figure A.3: The Effect of Public Health Units on Fertility

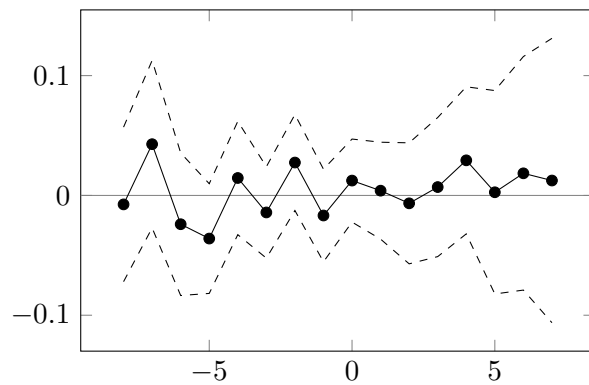
(a) Baseline Specification



(c) Dynamic Specification



(b) de Chaisemartin and D'Haultfoeuille (2020) Estimator



Notes: The dependent variables is the natural log of the ratio of births to population. Coefficients are weighted least-squares estimates. Dashed lines are 95 percent confidence intervals and are based on standard errors clustered by municipality. The year prior to the establishment of the PHU is omitted in Panel A, as is the first event-year in the sample in Panel B. Sample includes 76 municipalities and 1,672 municipality-year observations. In Panel A, event-years outside of the designated window are captured by “before” and “after” dummies.

#### A.1.4 PHU Dates of Inauguration

As discussed in Section 2.3, there are discrepancies in the inauguration dates for a few municipalities between official sources. In Section 2.5, PHU dates are based on the earliest opening date found in annual reports of the Commissioner of Health. In this section, I re-estimate Equation 2.1 to assess the sensitivity of the results to different ways of coding of  $PHU_{mk}$ . The pattern of results does not change, and the findings in this paper are not specific to the decisions made regarding PHU dates of inauguration.

Department of Health reports revised the year of inauguration to a later date for a few municipalities because earlier lists of PHUs had included semi-units, in which health officers were part-time employees, and units that did not provide rural coverage.<sup>1</sup> The discrepancies were due to the fact that in several towns “many changes had been made so that the skeleton of the new organization existed but the program was not complete...” (Commissioner of Health, 1932, p. 65). In any case, as seen in Tables A.1-A.4, the results for overall mortality, infant mortality, stillbirths, and TB mortality remain strong when semi-units are defined as untreated and municipalities without rural coverage are defined as untreated. Likewise, the results are robust to dates based on histories of PHUs written by a former director of PHUs (Chaves, 1942), which report later dates for semi-units and units without rural coverage, including when two obvious errors are fixed.<sup>2</sup> The same pattern of results also emerge when the inauguration date for Aguadilla is switched from 1931-32 to 1929-30 since Aguadilla had a health dispensary and was included in the list of municipalities with the skeleton of a PHU prior to 1932; and Loíza and Trujillo Alto are dropped, since they each briefly had a PHU, spent several years without a PHU, and later regained the PHU.

---

<sup>1</sup> The following municipalities were reported to have semi-units before 1931-32: Humacao, Juncos, and Mayagüez. The following municipalities were reported to have PHUs with no rural coverage before 1931-32: Humacao, Mayagüez, and Ponce. Additionally, San Juan was reported to have a PHU as early as 1928-29 (Governor of Puerto Rico, 1929) and Guayama in 1929-30 (Commissioner of Health, 1932), but later documents report 1931-32 as the year of inauguration for both municipalities. Guayama supposedly had a dispensary and unit organization nominally in place in 1929 but did not fully attempt its program until 1930.

<sup>2</sup> The PHUs in Manatí opened in 1931-32, not 1932-33 (Commissioner of Health, 1932), and in Sabana Grande in 1935-36, not 1934-35 (Commissioner of Health, 1936).

Table A.1: Robustness Checks: PHU Dates of Inauguration (Infant Mortality)

	(1)	(2)	(3)	(4)	(5)	(6)
Year -3	-0.02 (0.04)	0.00 (0.03)	0.00 (0.04)	0.00 (0.03)	-0.04 (0.03)	-0.00 (0.04)
Year -2	-0.01 (0.03)	-0.00 (0.03)	-0.00 (0.03)	-0.01 (0.03)	-0.02 (0.03)	-0.00 (0.03)
Year 0	-0.07** (0.03)	-0.06* (0.03)	-0.06* (0.03)	-0.06* (0.03)	-0.06** (0.03)	-0.07** (0.03)
Year 1	-0.05 (0.04)	-0.05 (0.04)	-0.05 (0.04)	-0.05 (0.04)	-0.05 (0.04)	-0.06 (0.04)
Year 2	-0.07** (0.04)	-0.07* (0.04)	-0.06 (0.04)	-0.06 (0.04)	-0.08** (0.04)	-0.07* (0.04)
Year 3	-0.09** (0.04)	-0.09** (0.04)	-0.09** (0.05)	-0.09** (0.05)	-0.10** (0.04)	-0.11** (0.05)
Year 4	-0.08 (0.05)	-0.09* (0.05)	-0.09 (0.05)	-0.08 (0.05)	-0.08* (0.05)	-0.10* (0.05)
Year 5	-0.11** (0.04)	-0.13*** (0.04)	-0.13** (0.05)	-0.13** (0.05)	-0.10** (0.04)	-0.15*** (0.05)
Year 6	-0.10** (0.05)	-0.12** (0.05)	-0.10* (0.05)	-0.11** (0.05)	-0.10** (0.04)	-0.13** (0.05)
Year 7	-0.15*** (0.05)	-0.16*** (0.05)	-0.14** (0.06)	-0.13** (0.06)	-0.14*** (0.05)	-0.16** (0.06)
PHU Dates	Semi-unit ≠ PHU	No rural → no PHU	Chaves	Chaves w/ 2 fixes	Aguadilla 1929-30	Drop Loíza, Truj. Alto
Obs.	1,672	1,672	1,672	1,672	1,672	1,628

Notes: The dependent variable is the log infant mortality rate. The coefficients are weighted least-squares estimates. The year prior to the establishment of the PHU is omitted. Event-years outside of the designated window are captured by “before” and “after” dummies. Standard errors are clustered by municipality. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table A.2: Robustness Checks: PHU Dates of Inauguration (Tuberculosis Mortality)

	(1)	(2)	(3)	(4)	(5)	(6)
Year -4	-0.05 (0.06)	-0.05 (0.06)	-0.05 (0.06)	-0.05 (0.06)	-0.05 (0.06)	-0.05 (0.06)
Year -3	-0.07 (0.05)	-0.07 (0.05)	-0.06 (0.06)	-0.06 (0.06)	-0.07 (0.05)	-0.07 (0.06)
Year -2	-0.03 (0.03)	-0.03 (0.03)	-0.03 (0.03)	-0.02 (0.03)	-0.04 (0.03)	-0.03 (0.03)
Year 0	-0.03 (0.03)	-0.03 (0.03)	-0.04 (0.04)	-0.04 (0.04)	-0.04 (0.03)	-0.05 (0.04)
Year 1	-0.10** (0.04)	-0.11** (0.05)	-0.13** (0.06)	-0.13** (0.06)	-0.12*** (0.04)	-0.13** (0.06)
Year 2	-0.14** (0.05)	-0.14** (0.06)	-0.16** (0.06)	-0.15** (0.06)	-0.15*** (0.05)	-0.16** (0.07)
Year 3	-0.14*** (0.05)	-0.14** (0.06)	-0.14** (0.07)	-0.13* (0.07)	-0.15*** (0.05)	-0.15** (0.07)
Year 4	-0.13* (0.07)	-0.12* (0.07)	-0.11 (0.08)	-0.11 (0.08)	-0.14** (0.06)	-0.13 (0.08)
Year 5	-0.14** (0.07)	-0.15** (0.07)	-0.15* (0.09)	-0.14 (0.09)	-0.17** (0.07)	-0.16* (0.09)
Year 6	-0.15* (0.09)	-0.14 (0.09)	-0.13 (0.10)	-0.13 (0.10)	-0.16* (0.08)	-0.15 (0.10)
Year 7	-0.20** (0.09)	-0.19* (0.10)	-0.18 (0.11)	-0.18 (0.11)	-0.22** (0.09)	-0.19 (0.11)
PHU Dates	Semi-unit ≠ PHU	No rural → no PHU	Chaves	Chaves w/ 2 fixes	Aguadilla 1929-30	Drop Loíza, Truj. Alto
Obs.	1,748	1,748	1,748	1,748	1,748	1,702

Notes: The dependent variable is the log TB mortality rate. The coefficients are weighted least-squares estimates. The year prior to the establishment of the PHU is omitted. Event-years outside of the designated window are captured by “before” and “after” dummies. Standard errors are clustered by municipality. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table A.3: Robustness Checks: PHU Dates of Inauguration (Overall Mortality)

	(1)	(2)	(3)	(4)	(5)	(6)
Year -4	0.00 (0.02)	-0.00 (0.02)	-0.02 (0.02)	-0.02 (0.02)	0.00 (0.02)	-0.03 (0.02)
Year -3	-0.02 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.02 (0.02)	-0.02 (0.02)
Year -2	-0.02 (0.02)	-0.01 (0.02)	-0.03* (0.02)	-0.03* (0.02)	-0.02 (0.02)	-0.03* (0.02)
Year 0	-0.03** (0.02)	-0.04** (0.02)	-0.04** (0.02)	-0.04** (0.02)	-0.03** (0.01)	-0.04** (0.02)
Year 1	-0.02 (0.02)	-0.03 (0.02)	-0.04* (0.02)	-0.04* (0.02)	-0.02 (0.02)	-0.04* (0.02)
Year 2	-0.04** (0.02)	-0.03** (0.02)	-0.03* (0.02)	-0.04** (0.02)	-0.05*** (0.02)	-0.04** (0.02)
Year 3	-0.06*** (0.02)	-0.05** (0.02)	-0.04** (0.02)	-0.04** (0.02)	-0.07*** (0.02)	-0.04** (0.02)
Year 4	-0.05* (0.03)	-0.05 (0.03)	-0.03 (0.03)	-0.03 (0.03)	-0.06** (0.03)	-0.04 (0.03)
Year 5	-0.04 (0.03)	-0.04 (0.03)	-0.04 (0.03)	-0.04 (0.03)	-0.05* (0.03)	-0.04 (0.03)
Year 6	-0.04 (0.03)	-0.04 (0.03)	-0.03 (0.04)	-0.03 (0.04)	-0.05 (0.03)	-0.03 (0.04)
Year 7	-0.06 (0.04)	-0.06 (0.04)	-0.05 (0.04)	-0.05 (0.04)	-0.07* (0.04)	-0.05 (0.04)
PHU Dates	Semi-unit ≠ PHU	No rural → no PHU	Chaves	Chaves w/ 2 fixes	Aguadilla 1929-30	Drop Loíza, Truj. Alto
Obs.	1,748	1,748	1,748	1,748	1,748	1,702

Notes: The dependent variable is the log mortality rate. The coefficients are weighted least-squares estimates. The year prior to the establishment of the PHU is omitted. Event-years outside of the designated window are captured by “before” and “after” dummies. Standard errors are clustered by municipality. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table A.4: Robustness Checks: PHU Dates of Inauguration (Stillbirths)

	(1)	(2)	(3)	(4)	(5)	(6)
Year -3	0.02 (0.05)	0.03 (0.05)	0.05 (0.06)	0.06 (0.06)	0.00 (0.04)	0.01 (0.05)
Year -2	-0.01 (0.05)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)	-0.03 (0.05)	-0.01 (0.04)
Year 0	0.01 (0.04)	0.02 (0.03)	0.00 (0.03)	0.00 (0.03)	-0.01 (0.04)	0.00 (0.03)
Year 1	-0.03 (0.04)	-0.04 (0.05)	-0.05 (0.05)	-0.04 (0.05)	-0.04 (0.04)	-0.04 (0.05)
Year 2	-0.02 (0.06)	-0.04 (0.06)	-0.06 (0.07)	-0.06 (0.07)	-0.02 (0.06)	-0.03 (0.06)
Year 3	-0.08 (0.08)	-0.09 (0.08)	-0.14 (0.09)	-0.13 (0.09)	-0.09 (0.08)	-0.10 (0.08)
Year 4	-0.08 (0.09)	-0.09 (0.09)	-0.12 (0.10)	-0.13 (0.10)	-0.09 (0.08)	-0.07 (0.08)
Year 5	-0.08 (0.10)	-0.08 (0.11)	-0.12 (0.12)	-0.13 (0.12)	-0.08 (0.10)	-0.05 (0.10)
Year 6	-0.10 (0.12)	-0.10 (0.12)	-0.12 (0.14)	-0.13 (0.14)	-0.11 (0.11)	-0.04 (0.11)
Year 7	-0.10 (0.13)	-0.09 (0.13)	-0.14 (0.15)	-0.15 (0.15)	-0.11 (0.12)	-0.05 (0.12)
PHU Dates	Semi-unit ≠ PHU	No rural → no PHU	Chaves	Chaves w/ 2 fixes	Aguadilla 1929-30	Drop Loíza, Truj. Alto
Obs.	1,672	1,672	1,672	1,672	1,672	1,628

Notes: The dependent variable is the log stillbirth rate. The coefficients are weighted least-squares estimates. The year prior to the establishment of the PHU is omitted. Event-years outside of the designated window are captured by “before” and “after” dummies. Standard errors are clustered by municipality. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

### A.1.5 Municipality-Specific Time Trends

A common practice in analyses using an event study framework is to relax the assumption of common trends by allowing for unit-specific time trends. Following the guidance of Goodman-Bacon (forthcoming) and Borusyak and Jaravel (2018), I do not include municipality-specific time trends in the main analysis. Unit-specific linear trends tend to absorb time-varying treatment effects that are necessarily larger at the end of the panel, and in these cases they over control. Nonetheless, such trends are included as a robustness check, and the main results are unaffected (see Tables A.5-A.8).

### A.1.6 Census Controls

Time-varying municipal characteristics were not frequently reported throughout the sample period. In this section, I extract several variables from the the 1920, 1930, 1940, and 1950 censuses of Puerto Rico and linearly impute intercensal years. Following Anderson et al. (2019a), these variables are the fraction of the population that was female, nonwhite, younger than 5, and literate.<sup>3</sup> The fraction of the population that was under 5 is endogenous to infant mortality, and is significantly negatively correlated with changes in the infant mortality rate. The fractions female (due to the prevalence of puerperal mortality) or literate (due to changes in incentives to invest in human capital) may also be endogenous to mortality. In any case, results for Equation 2.1 appear in Tables A.5-A.8 and demonstrate, once again, that PHUs significantly reduced overall, TB, and infant mortality, and may have reduced stillbirths.

---

<sup>3</sup> Anderson et al. (2019a) uses the fraction of the population younger than 18, but I use the fraction under 5 because it is reported in the census report for each of the four censuses.

Table A.5: Robustness Checks: Alternative Specifications (Infant Mortality)

	(1)	(2)	(3)	(4)
Year -3	-0.03 (0.03)	-0.03 (0.03)	-0.03 (0.03)	0.00 (0.03)
Year -2	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)
Year 0	-0.06** (0.03)	-0.06** (0.03)	-0.06* (0.03)	-0.06* (0.03)
Year 1	-0.04 (0.03)	-0.04 (0.03)	-0.04 (0.04)	-0.05 (0.04)
Year 2	-0.08** (0.04)	-0.07** (0.04)	-0.07** (0.04)	-0.09* (0.05)
Year 3	-0.10** (0.04)	-0.09** (0.04)	-0.09** (0.04)	-0.13** (0.05)
Year 4	-0.08* (0.04)	-0.07 (0.05)	-0.08 (0.05)	-0.11* (0.06)
Year 5	-0.10** (0.04)	-0.09** (0.04)	-0.10** (0.04)	-0.17*** (0.05)
Year 6	-0.10** (0.05)	-0.10** (0.05)	-0.10** (0.05)	-0.14** (0.06)
Year 7	-0.14*** (0.05)	-0.14** (0.06)	-0.14*** (0.05)	-0.20*** (0.06)
Robustness Check	Muni- trends	Census Controls	Public Water	Un- weighted

Notes: Notes: The dependent variable is the log stillbirth rate. Observations are weighted by municipal population in columns 1 to 3. The year prior to the establishment of the PHU is omitted. Event-years outside of the designated window are captured by “before” and “after” dummies. All regressions include 1,672 observations. Column 1 includes municipality-specific linear trends. Column 2 includes controls for the fraction of the population that was female, nonwhite, younger than 5, and literate, linearly interpolated from the 1920, 1930, 1940, and 1950 censuses. Column 3 includes an indicator variable equal to one if municipality  $m$  has a public water source in year  $t$  and zero otherwise. Column 4 is unweighted. Standard errors are clustered by municipality. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table A.6: Robustness Checks: Alternative Specifications (Tuberculosis Mortality)

	(1)	(2)	(3)	(4)
Year -4	-0.07 (0.06)	-0.05 (0.06)	-0.04 (0.06)	-0.11 (0.09)
Year -3	-0.09 (0.06)	-0.08 (0.06)	-0.07 (0.05)	-0.17* (0.10)
Year -2	-0.04 (0.03)	-0.04 (0.03)	-0.03 (0.03)	-0.06* (0.04)
Year 0	-0.03 (0.03)	-0.03 (0.03)	-0.03 (0.03)	-0.06 (0.04)
Year 1	-0.11** (0.04)	-0.11** (0.04)	-0.12*** (0.04)	-0.16** (0.06)
Year 2	-0.14** (0.06)	-0.15*** (0.05)	-0.16*** (0.05)	-0.19** (0.08)
Year 3	-0.13** (0.06)	-0.14*** (0.05)	-0.16*** (0.05)	-0.15** (0.06)
Year 4	-0.12 (0.08)	-0.13* (0.06)	-0.14** (0.07)	-0.11 (0.07)
Year 5	-0.14 (0.09)	-0.15** (0.07)	-0.17** (0.07)	-0.17** (0.08)
Year 6	-0.14 (0.11)	-0.14* (0.08)	-0.16* (0.08)	-0.14 (0.10)
Year 7	-0.19 (0.12)	-0.19** (0.09)	-0.22** (0.09)	-0.23** (0.11)
Robustness Check	Muni-trends	Census Controls	Public Water	Un-weighted

Notes: The dependent variable is the log stillbirth rate. Observations are weighted by municipal population in columns 1 to 3. The year prior to the establishment of the PHU is omitted. Event-years outside of the designated window are captured by “before” and “after” dummies. All regressions include 1,748 observations. Column 1 includes municipality-specific linear trends. Column 2 includes controls for the fraction of the population that was female, nonwhite, younger than 5, and literate, linearly interpolated from the 1920, 1930, 1940, and 1950 censuses. Column 3 includes an indicator variable equal to one if municipality  $m$  has a public water source in year  $t$  and zero otherwise. Column 4 is unweighted. Standard errors are clustered by municipality. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table A.7: Robustness Checks: Alternative Specifications (Overall Mortality)

	(1)	(2)	(3)	(4)
Year -4	0.01 (0.02)	-0.00 (0.02)	0.00 (0.02)	-0.01 (0.02)
Year -3	-0.02 (0.02)	-0.03 (0.02)	-0.02 (0.02)	-0.02 (0.02)
Year -2	-0.01 (0.02)	-0.02 (0.02)	-0.02 (0.02)	-0.04** (0.02)
Year 0	-0.03** (0.01)	-0.03** (0.01)	-0.03** (0.01)	-0.04** (0.02)
Year 1	-0.02 (0.02)	-0.02 (0.02)	-0.02 (0.02)	-0.02 (0.02)
Year 2	-0.05*** (0.02)	-0.05*** (0.02)	-0.05*** (0.02)	-0.04** (0.02)
Year 3	-0.07*** (0.02)	-0.06*** (0.02)	-0.07*** (0.02)	-0.05** (0.02)
Year 4	-0.06** (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.04 (0.03)
Year 5	-0.06** (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.04 (0.03)
Year 6	-0.05 (0.03)	-0.04 (0.03)	-0.05 (0.03)	-0.03 (0.04)
Year 7	-0.08* (0.04)	-0.07* (0.04)	-0.07* (0.04)	-0.06 (0.04)
Robustness Check	Muni- trends	Census Controls	Public Water	Un- weighted

Notes: The dependent variable is the log stillbirth rate. Observations are weighted by municipal population in columns 1 to 3. The year prior to the establishment of the PHU is omitted. Event-years outside of the designated window are captured by “before” and “after” dummies. All regressions include 1,748 observations. Column 1 includes municipality-specific linear trends. Column 2 includes controls for the fraction of the population that was female, nonwhite, younger than 5, and literate, linearly interpolated from the 1920, 1930, 1940, and 1950 censuses. Column 3 includes an indicator variable equal to one if municipality  $m$  has a public water source in year  $t$  and zero otherwise. Column 4 is unweighted. Standard errors are clustered by municipality. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table A.8: Robustness Checks: Alternative Specifications (Stillbirths)

	(1)	(2)	(3)	(4)
Year -3	0.01 (0.05)	-0.00 (0.04)	-0.00 (0.04)	0.05 (0.07)
Year -2	-0.02 (0.05)	-0.02 (0.05)	-0.03 (0.05)	0.03 (0.05)
Year 0	-0.01 (0.04)	-0.01 (0.04)	-0.01 (0.04)	-0.00 (0.05)
Year 1	-0.06 (0.04)	-0.04 (0.05)	-0.04 (0.04)	-0.06 (0.06)
Year 2	-0.05 (0.06)	-0.02 (0.06)	-0.03 (0.06)	-0.07 (0.08)
Year 3	-0.10 (0.07)	-0.07 (0.08)	-0.08 (0.08)	-0.13 (0.11)
Year 4	-0.10 (0.07)	-0.07 (0.08)	-0.08 (0.08)	-0.14 (0.13)
Year 5	-0.09 (0.08)	-0.07 (0.10)	-0.08 (0.10)	-0.15 (0.15)
Year 6	-0.11 (0.09)	-0.10 (0.11)	-0.10 (0.11)	-0.16 (0.18)
Year 7	-0.10 (0.10)	-0.09 (0.12)	-0.10 (0.12)	-0.19 (0.20)
Robustness Check	Muni-trends	Census Controls	Public Water	Un-weighted

Notes: The dependent variable is the log stillbirth rate. Observations are weighted by municipal population in columns 1 to 3. The year prior to the establishment of the PHU is omitted. Event-years outside of the designated window are captured by “before” and “after” dummies. All regressions include 1,672 observations. Column 1 includes municipality-specific linear trends. Column 2 includes controls for the fraction of the population that was female, nonwhite, younger than 5, and literate, linearly interpolated from the 1920, 1930, 1940, and 1950 censuses. Column 3 includes an indicator variable equal to one if municipality  $m$  has a public water source in year  $t$  and zero otherwise. Column 4 is unweighted. Standard errors are clustered by municipality. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

### **A.1.7 Controlling for a Public Water System**

Improvements in water purification played a crucial role in the mortality transition in the United States (Alsan and Goldin, 2019; Anderson et al., 2019b, 2020b; Cutler and Miller, 2005; Ferrie and Troesken, 2008). Puerto Rico likewise invested in centralized water purification in the early 20th century, including during the period of analysis in this paper. Although the work of public health units involved the enforcement of sanitary regulations, PHUs were not responsible for the provision of drinking water. In Tables A.5-A.8, I control for improvements in drinking water by including an indicator variable equal to one if a municipality has a public water source and zero otherwise, derived from years of construction as reported in the 1939-40 annual report of the commissioner of health. The coefficient on the public water dummy (not shown) is negative for all forms of mortality but is never close to statistically significant, and its inclusion does not change any of the main results.

### **A.1.8 Unweighted Observations**

All regressions up to this point weighted observations by municipality population in 1925, generating results which generalize to the island population. In this section, I re-estimate Equation 2.1 without weighting observations to determine whether the main results were driven by a few large municipalities. The results are similar to those from the weighted regressions, indicating that PHUs played a major role in reducing all measured forms of mortality other than malaria mortality (see Tables A.5-A.8). In the case of overall mortality, there is some evidence of a significant change in mortality preceding PHU inauguration, as the point estimate on the lead variable two years prior equals -0.04 and is statistically significant at the 5 percent level. There is also some concern about differential pre-trends in the case of TB mortality, as the coefficients are fairly large and are significant at the 10 percent level in event-years -3 and -2. Still, the body of evidence supports the conclusion that PHUs reduced mortality across municipalities.

### A.1.9 Spillovers between Municipalities

Although PHUs were intended to improve local coordination of public health efforts at the municipal level, it is possible that PHUs could have had spillover effects on mortality in neighboring municipalities. Knowledge about infant hygiene, for example, may have diffused informally across society. None of the government documents reviewed for this study mentioned that services were restricted to inhabitants of the municipality in which the PHU was located, so PHUs may have also directly served inhabitants of nearby municipalities. In the case of TB, the reduction in morbidity in one municipality may reduce morbidity in nearby municipalities by slowing the spread of disease.

Table A.9: The Effect of PHUs on Mortality Rates Controlling for Neighboring Units

<i>Dependent variables:</i>	IMR	TB	Mortality	Stillbirths
PHU	-0.09** (0.03)	-0.16*** (0.05)	-0.05** (0.02)	-0.07 (0.08)
Neighboring PHU	0.00 (0.01)	0.10** (0.04)	0.00 (0.01)	-0.01 (0.05)
Observations	1,650	1,725	1,725	1,650

Notes: The dependent variables are log mortality rates. The coefficients are weighted least-squares estimates. Standard errors are clustered by municipality. The coefficient on *PHU* is derived from a linear combination of coefficients estimated separately for event-years 0 to 7. *NeighboringPHU<sub>mt</sub>* equals 1 if there is a PHU in a municipality bordering municipality *m* in year *t* or 0 otherwise. See Section A.1.9 for more detail. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table A.9 reports results from controlling for the opening of a dispensary in an adjacent municipality with an indicator variable equal to 1 if there is a PHU in a neighboring municipality or 0 otherwise. There is no evidence to suggest that PHUs had spillover effects into neighboring municipalities for any of the outcome variables. The point estimates for the effect of a PHU in an adjacent municipality are close to zero for infant mortality, general mortality, and stillbirths. This is not surprising since modern forms of transportation were still inaccessible for most families in this

era, so PHUs likely served the inhabitants of their municipality exclusively.<sup>4</sup> For TB mortality, the estimate is statistically significant but positive. The positive association between TB mortality and PHUs in neighboring municipalities is not robust to estimating the relationship flexibly in an event study, however.

#### A.1.10 Spatial Autocorrelation

Spatial data tend to be autocorrelated, and  $t$ -statistics may be too large in regressions with spatially autocorrelated residuals (Kelly, 2019). Clustering standard errors is invalid in the presence of non-negligible spatial autocorrelation since residuals in neighboring clusters will tend to be correlated. The regression results based on clustering at the municipal level do not account for spatial autocorrelation.

Table A.10 presents results for Equation 2.1 with standard errors adjusted for arbitrary dependence of the errors across observations in space and across time periods as in Colella et al. (2019), building on Conley (1999). I present results for a distance cutoff of 30 km. and that allow for dependence across all time periods. These results are conservative in that different parameters generally produce smaller standard errors. Still, the standard errors from this procedure are smaller than the clustered standard errors presented in the body of the paper. All of the main results of the paper hold after adjusting for spatial autocorrelation.

---

<sup>4</sup> For instance, nearly one-half of rural homes in the interior of the island were located more than three miles from a paved or dirt road (Mountin et al., 1937).

Table A.10: The Effect of Public Health Units on Mortality Rates by Cause of Death (Standard Errors Adjusted for Spatial Autocorrelation)

<i>Dependent variable:</i>	IMR	TB	Mortality	Stillbirths
	(1)	(2)	(3)	(4)
Year -4		-0.05 (0.06)	0.00 (0.02)	
Year -3	-0.03 (0.03)	-0.07 (0.06)	-0.02 (0.02)	-0.00 (0.04)
Year -2	-0.01 (0.03)	-0.03 (0.03)	-0.02 (0.02)	-0.03 (0.05)
Year 0	-0.06** (0.03)	-0.03 (0.04)	-0.03** (0.01)	-0.01 (0.03)
Year 1	-0.04 (0.03)	-0.12** (0.05)	-0.02 (0.01)	-0.05 (0.04)
Year 2	-0.07** (0.03)	-0.16*** (0.06)	-0.05*** (0.02)	-0.03 (0.06)
Year 3	-0.09** (0.04)	-0.15*** (0.05)	-0.06*** (0.02)	-0.09 (0.08)
Year 4	-0.07* (0.04)	-0.14** (0.06)	-0.05** (0.03)	-0.09 (0.08)
Year 5	-0.10** (0.04)	-0.16** (0.06)	-0.05* (0.03)	-0.08 (0.10)
Year 6	-0.10** (0.04)	-0.16** (0.08)	-0.05 (0.03)	-0.11 (0.11)
Year 7	-0.14*** (0.05)	-0.21** (0.08)	-0.07* (0.04)	-0.11 (0.13)
Observations	1,672	1,748	1,748	1,672

Notes: The dependent variables are log mortality rates. The coefficients are weighted least-squares estimates. The year prior to the establishment of the PHU is omitted. Event-years outside of the designated window are captured by “before” and “after” dummies. Standard errors are adjusted for spatial autocorrelation following Colella et al. (2019) with a distance cutoff of 30 km. and a time lag covering the entire study period. See Appendix A.1.10 for more details. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

### A.1.11 Static Difference-in-Differences

In difference-in-differences (DiD) settings with staggered adoption (e.g. the rollout of PHUs in Puerto Rico), the DiD coefficient in two-way (unit and time) fixed effects models does not have a clear causal interpretation under treatment effect heterogeneity. The DiD coefficient represents a weighted average of dynamic effects where some of the weights can be negative; in extreme cases, the DiD coefficient could be negative (positive) even though the treatment effect on each cohort is positive (negative) (Abraham and Sun, forthcoming; Borusyak and Jaravel, 2018; de Chaisemartin and D’Haultfoeuille, 2020; Goodman-Bacon, forthcoming). Using the `twowayfweights` Stata package in a DiD model under the common trends assumption, I find that 32 percent of weights are negative; the DiD coefficient does not represent an average treatment effect and is uninterpretable (de Chaisemartin and D’Haultfoeuille, 2020).

Still, “static” DiD models, with a treatment dummy instead of leads and lags around the treatment event, are common in social science research. For comparison, I present results from a static model in the first row of Table A.11 while stressing that these results are difficult to interpret and do not summarize the treatment effect. The PHU indicator variable is lagged one period when the stillbirth rate or the TB mortality rate are the dependent variable since the event study revealed that the effect on these outcomes was delayed. Across the board the effect is negative, and it is statistically significant at the 5 percent level for TB and the 10 percent level for general and malaria mortality.

In the second and third rows of Table A.11, I provide alternative “static” DiD coefficients by manually averaging the coefficients from the baseline (“capped”) and fully dynamic specifications. The coefficient derived from the dynamic specification, with a full set of event-year dummies (with the first event-year and event-year -1 omitted) has the advantages of summarizing the treatment effects for the first eight years with a PHU while avoiding the problem of negative weighting (from dynamic treatment effects only) by estimating treatment effects flexibly. The dynamic model does not estimate treatment effects efficiently, however. I also estimate a model with the pre-period

binned but with a full series of treatment indicators.

The baseline model suggests that PHUs significantly reduced TB mortality at the 1 percent level and infant and general mortality at the 5 percent level. The dynamic specification suggests that PHUs significantly reduced infant mortality at the 1 percent level. The coefficients are negative and quite large across specifications and outcome variables, strengthening the conclusion that PHUs significantly reduced these forms of mortality. The coefficients on malaria mortality are consistently negative but only statistically significant, or close to significance, in the static DiD specification, which is difficult to interpret. In any case, as demonstrated in Section 2.5, the disaggregated event study results do not indicate that PHUs had an effect on malaria mortality.

Table A.11: The Effect of Public Health Units on Mortality Rates (Difference-in-Differences Specifications)

<i>Dependent variable:</i>	IMR	TB	Mortality	Stillbirths	Malaria
	(1)	(2)	(3)	(4)	(5)
Static DiD	-0.02 (0.03)	-0.09** (0.03)	-0.03* (0.01)	-0.04 (0.04)	-0.16* (0.10)
Baseline Specification	-0.08** (0.03)	-0.14*** (0.05)	-0.05** (0.02)	-0.07 (0.08)	-0.13 (0.16)
Dynamic Specification	-0.19*** (0.06)	-0.11 (0.07)	-0.06 (0.04)	-0.08 (0.07)	-0.27 (0.50)
Binned Pre-period, Dynamic Specification	-0.12*** (0.04)	-0.10 (0.07)	-0.05** (0.02)	-0.19 (0.12)	-0.02 (0.21)

Notes: The dependent variables are log mortality rates. The coefficients are weighted least-squares estimates. Standard errors are clustered by municipality. The coefficient shown in the first row is based on an indicator variable equal to 1 if a municipality is served by a PHU in that year. The coefficients from the baseline and dynamic specifications are linear combinations of dynamic event study treatment effects from Year 0 to Year 7. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

### A.1.12 Infant Mortality, Maternal Mortality, and Midwives (Alternative Specification)

Results based on a difference-in-differences model in Section 2.5.3 indicate that infant and maternal mortality are negatively associated with assistant midwives per capita starting in 1931. In this section I test the robustness of that result with a generalized difference-in-difference model to determine when these negative relationships emerged.

Figure A.4 displays results from a generalized difference-in-difference model which includes a series of year fixed effects multiplied by the number of assistant midwives per 1,000 inhabitants in 1935. The interaction term for 1924 is omitted. The pattern of coefficients over time suggests that the negative association between midwives per capita and infant mortality emerged in fiscal year 1929-30, just prior to the 1931 law mandating that assistant midwives be licensed. Coefficients trend downwards again in 1932 after briefly moving back up around zero. Still, the negative association before 1931 raises concern that the relationship between midwives per capita and infant mortality is driven by an omitted variable. However, in the second row of the first column, midwives per capita is replaced with the share of births attended by authorized persons—a proxy for the the impact of midwives, since midwives delivered 85 percent of infants—and a negative relationship between authorized births and infant mortality arises after the 1931 law. Hence, the evidence is suggestive, but not conclusive, that infant mortality is related to the licensing of midwives in 1931.

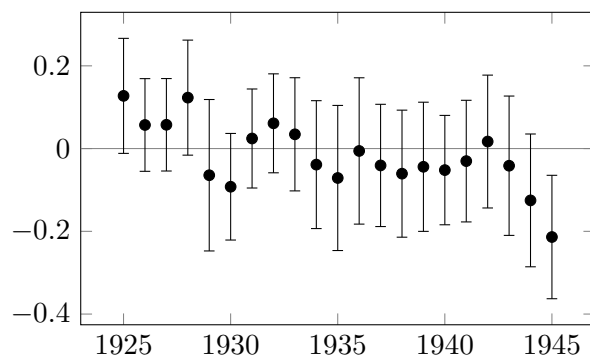
The right column of Figure A.4 displays results for maternal mortality. The pattern of coefficients suggests that the negative association between midwives per capita and maternal mortality did not emerge until 1931, although the two years preceding the law are missing. The negative association remains through the rest of the sample period but increases in magnitude in 1937 and subsequent years, around the time when sulfa drugs became widely available and reduced maternal mortality (Jayachandran et al., 2010; Thomasson and Treber, 2008). While midwives probably did not administer sulfa drugs, midwives might have recognized abnormal symptoms and known to call the attention of a physician. In any case, results for 1937 to 1945 do not drive the

Figure A.4: Treatment Heterogeneity by Midwives Per Capita  
(Infant and Maternal Mortality)

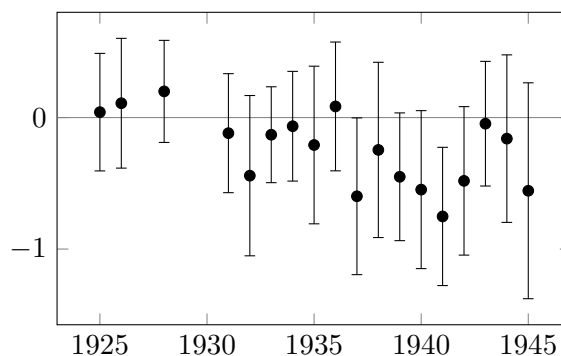
(a) Infant Mortality

(b) Maternal Mortality

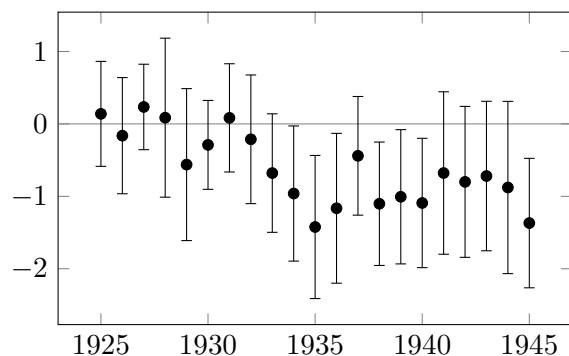
Midwives per capita



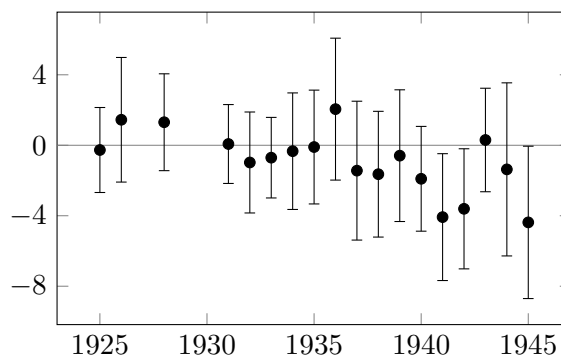
Midwives per capita



Authorized deliveries



Authorized deliveries



Notes: The dependent variables are log mortality rates. Coefficients are weighted least-squares estimates based on a generalized difference-in-differences model that includes year and municipality fixed effects and a full set of year fixed effects interacted with the number of assistant midwives per capita in 1935 (top row) or the share of births attended by authorized persons (i.e. assistant midwives, nurse midwives, or physicians) in 1935-36 (bottom row), with the first year of data omitted. Errors bars are 95 percent confidence intervals and are based on standard errors clustered by municipality. Sample includes 76 municipalities and 1,672 municipality-year observations for infant mortality and 1,444 municipality-year observations for maternal mortality.

significant correlation between midwives per capita and maternal mortality: excluding later years in the difference-in-differences model, an additional midwife per 1,000 inhabitants is associated with 23 percent lower maternal mortality ( $p = 0.035$ ). Results based on the share of births attended by authorized persons follow a similar pattern, albeit weaker.

## A.2 Heterogeneity by Socioeconomic Characteristics

The diversity of municipalities motivated the decentralization of public health, recognizing that officials in San Juan were not equipped to identify and respond to challenges elsewhere on the island. For instance, the adult literacy rate in 1920 in the capital and primary urban center, San Juan, stood at 74 percent, while in the town of Cidra, located in the rural, mountainous interior, the rate was just 28 percent. More relevant to this investigation, municipalities differed in their exposure to disease and access to professional health care: tuberculosis was nearly twice as high in urban as in rural areas (Rodríguez Pastor and Janer, 1953), and more than half of all physicians were located in just five municipalities (Mountin et al., 1937). Given such diversity, and with coordination of policy at the local level, the effect of PHUs might be expected to vary across municipalities.

Table A.12 provides evidence of treatment heterogeneity based on a model with lags interacted with time-invariant municipal socioeconomic characteristics. Reported coefficients are linear combinations of coefficients from event-years 0 through 7. PHUs were more effective in municipalities with higher pre-treatment mortality rates for each of the four outcomes, implying convergence in mortality, perhaps as a result of convergence in hygienic knowledge and access to health services. Larger effects in areas with higher mortality parallels, e.g., Bailey and Goodman-Bacon (2015) and Hansen et al. (2020).

PHUs more successfully reduced the burden of TB in municipalities with more physicians per capita, lower shares of agricultural employment, greater shares of the population in urban areas, and higher literacy rates. This result is not surprising since TB mortality was much higher in urban areas and anti-TB efforts were concentrated in the largest cities and towns.<sup>5</sup> There is weak evidence that PHUs more successfully reduced infant mortality in more rural areas, as the coefficient on the interaction variable is just shy of statistical significance at the 10 percent level for doctors per

---

<sup>5</sup> For example, the anti-TB centers, which handled diagnoses and pneumothorax treatment, were located in the following municipalities: San Juan (population rank in 1935: 1), Ponce (2), Mayagüez (3), Arecibo (4), Caguas (6), Bayamón (8), Aguadilla (9), Guayama (15), and Fajardo (36) (Commissioner of Health, 1938).

capita, share of the labor force in agriculture, urban share, and literacy. Prenatal and postpartum clinics at the health units were targeted at poor and indigent mothers (Commissioner of Health, 1942), and dozens of rural dispensaries opened by the Puerto Rico Reconstruction Administration in 1936 were subsumed by PHUs in 1938 (Ramírez de Arellano, 1981). The steeper decline in infant mortality in rural areas comports with Hoehn-Velasco (2018), who finds that county health departments were more effective at reducing infant mortality in rural counties in the United States in the early 20th century.

Table A.12: Treatment Heterogeneity of Public Health Units by Municipal Characteristics

<i>Dependent variable:</i>	IMR	TB	Mortality	Stillbirths
	(1)	(2)	(3)	(4)
PHU	-0.07 (0.05)	0.04 (0.06)	-0.02 (0.03)	-0.06 (0.10)
PHU × $Y_{1923-26}$	-0.11*** (0.02)	-0.12*** (0.03)	-0.05*** (0.01)	-0.18*** (0.04)
PHU	-0.15*** (0.05)	0.00 (0.05)	-0.04 (0.03)	-0.20 (0.13)
PHU × Physicians	0.02 (0.01)	-0.07*** (0.01)	-0.01 (0.01)	0.00 (0.01)
PHU	-0.16*** (0.06)	-0.01 (0.06)	-0.04 (0.03)	-0.15 (0.12)
PHU × Agricultural Employment	-0.03 (0.02)	0.07** (0.03)	0.00 (0.01)	0.04 (0.03)
PHU	-0.16** (0.06)	0.00 (0.06)	-0.03 (0.03)	-0.21 (0.14)
PHU × Urban share	0.02 (0.02)	-0.07*** (0.02)	-0.01 (0.01)	0.01 (0.03)
PHU	-0.16*** (0.05)	-0.02 (0.06)	-0.04 (0.03)	-0.18 (0.13)
PHU × Literacy	0.03 (0.02)	-0.07*** (0.02)	-0.01 (0.01)	-0.01 (0.02)

Notes: The dependent variables are log mortality rates. The coefficients are weighted least-squares estimates. Standard errors are clustered by municipality.  $Y_{1923-26}$  represents the average value of the dependent variable (not in log form) for municipality  $m$  from 1924 to 1926.  $Physicians_s_m$  is the number of physicians (public or private) serving  $m$  in 1935.  $AgriculturalEmployment_m$  is the percent of the labor force in  $m$  employed in agriculture in 1935.  $UrbanShare_m$  is the share of the population of  $m$  in 1930 living in urban areas (defined as towns with at least 2,500 inhabitants).  $Literacy_m$  is the share of the adult population literate in 1920. Each cross-section of municipality characteristics is standardized to have a mean equal to zero and a standard deviation equal to one. Coefficients on municipal characteristics are estimated as a full set of lags interacted with the municipal characteristic. The coefficients reported are a linear combination of coefficients estimated separately for event-years 0 to 7. Horizontal lines delineate separate series of regressions. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

### A.3 Additional Tables

Table A.13: Selected Activities of Public Health Units in Fiscal Year 1938-39

Public Health Education:	
Health Talks .....	25,873
Attendance .....	717,057
Home Visits Made by Nurses .....	164,623
Persons Receiving Specific Immunization or Treatment:	
Vaccination Against Smallpox .....	48,162
Immunization Against Typhoid Fever .....	44,252
Immunization Against Diphtheria .....	8,247
Treatment for Uncinariasis .....	207,640
Clinical Consultations:	
Tuberculosis Clinics .....	6,968
Prenatal Clinics .....	52,718
Infant Hygiene Clinics .....	61,035
Preschool Clinics .....	28,323
Number of School Children Examined .....	170,576

Source: *Report of the Commissioner of Health of Puerto Rico for the Fiscal Year 1938-39*

Table A.14: Leading Causes of Death, 1931-32

(a) All deaths

Causes of Death	Death Rate per 100,000 population	Percent of all deaths
1. Diarrhea and Enteritis	408.2	18.3
2. Tuberculosis	297.3	13.4
3. Pneumonia	232.5	10.4
4. Malaria	174.9	7.8
5. Nephritis	137.4	6.2

(b) Deaths under one year of age

Causes of Death	Death Rate per 1,000 live births	Percent of all deaths
1. Diarrhea and Enteritis	39.6	29.9
2. Pneumonia	15.3	11.6
3. Bronchitis	7.9	6.0
4. Tetanus	5.8	4.4
5. Malaria	5.8	4.4

Source: *Thirty-Second Annual Report of the Governor of Puerto Rico*

Table A.15: Mortality Rates per 100,000 Inhabitants for the Principal Infectious and Parasitic Diseases, 1931-1949

Years	Diarrhea/ Enteritis	TB	Malaria	Pneu- monia	Bron- chitis	Uncin- ariasis	Influenza	Tetanus	Total
1931-34	388.0	300.7	180.3	188.4	51.2	33.9	29.8	21.1	1193.4
1935-39	426.4	280.2	117.0	169.7	41.6	16.6	24.1	8.1	1083.6
1940-44	348.0	235.6	85.8	142.5	23.4	12.5	23.0	7.3	879.2
1945-49	201.7	185.9	21.9	108.7	13.7	1.9	10.1	6.3	550.3

Source: Vázquez Calzada (1978)

## Appendix B

### Supplemental Material for Chapter 3

#### B.1 Data

##### B.1.1 Dependent Variables

- (1) **Human Built-up and Settlement (Urban) Extent, 2010:** high spatial resolution estimates of global urban extent, derived from the Global Land Survey Landsat dataset for the target year 2010.<sup>1</sup>
- (2) **Nighttime Light Density, 2012:** derived from cloud-free composites of nighttime light collected from satellites by the US Air Force Weather Agency and processed by the National Oceanic and Atmospheric Administration's National Geophysical Data Center. The dataset used is the F15 2003 Nighttime Lights Composite from the Version 4 DMSP-OLS Nighttime Lights Time Series.<sup>2</sup>
- (3) **Number of Corporations, 2017:** total number of corporations which listed a given municipality as its physical city in the *Registro de Corporaciones*. The dataset includes all corporations currently active and that participated in the most recent collection of economic data. The data were last updated on March 13, 2017 and are provided by the Puerto Rico State Department.<sup>3</sup>
- (4) **Number of Factories, 1966:** total number of factories established under the development program Operation Bootstrap (*Operación Manos a la Obra* in Spanish) in a given

---

<sup>1</sup> Available at <http://sedac.ciesin.columbia.edu/data/set/ulandsat-hbase-v1>. Last accessed 2/4/2019.

<sup>2</sup> Available at <http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>. Last accessed 2/4/2019.

<sup>3</sup> Available at <http://data.pr.gov/en/data>. Last accessed 2/4/2019.

municipality as of 1966. The source of the information is the Puerto Rico Economic Development Administration, and the data were transcribed from Picó (1969, p. 327).

- (5) **Per Capita Income, 2012-2016:** the mean income computed for every man, woman, and child in a municipality, averaged over the years 2012 to 2016. The data are reported by the US Census Bureau.<sup>4</sup>
- (6) **Population Density:** population divided by area, calculated at the municipal level using population counts from the US Census; calculated for 4.5 km.  $\times$  4.5 km. squares using *barrio*-level data from the US Census.
- (7) **Index of Municipal Socioeconomic Development, 2013:** produced by the Puerto Rico Planning Board based on per capita income, median family income, families below the poverty level, unemployment rate, population education level, and illiteracy. The index is modeled on the United Nations Human Development Index.<sup>5</sup>

### B.1.2 Development Controls

- (1) **Distance to port:** distance from the centroid of a grid cell or municipality to the nearest port of entry. I include ports of entry as reported by the Department of Commerce and Labor (1907): Aguadilla, Arecibo, Arroyo, Fajardo, Guánica, Humacao, Mayagüez, Ponce, and San Juan.
- (2) **Distance to San Juan:** distance from the centroid of a grid cell or municipality to San Juan.
- (3) **Percentage literate, 1899:** the number of residents ages 10 and up who can read divided by the total population over the age of 10, as reported by the 1899 Census.
- (4) **Percentage of labor in agriculture, 1899:** The number of workers, male or female,

---

<sup>4</sup> Available at <https://www.census.gov/quickfacts/>. Last accessed 2/4/2019.

<sup>5</sup> Available at <https://data.pr.gov/>. Last accessed 2/4/2019.

employed in agriculture, divided by the number of workers in the labor force, as reported by the 1899 Census.

- (5) **Population density, 1899:** population divided by area. For the gridded dataset, population density is calculated using *barrio*- (district-)level population counts which are averaged over the area of each  $4.5 \times 4.5$  km. square.

### B.1.3 Transportation Data

- (1) **Distance to roads:** distance from the centroid of a cell to the nearest highway, cart road, or railroad in 1899, 1920, or 1940, respectively. I used maps from George F. Cram<sup>6</sup> and the US War Department for 1899,<sup>7</sup> Rand McNally for 1920,<sup>8</sup> and the Puerto Rico Bureau of Public Works for 1940.<sup>9</sup> I have also consulted several descriptive sources to classify road quality and cross-check accuracy of maps, including the War Department (1898), Governor of Puerto Rico (1901), Carroll (1899), Robinson (1899), and Wilson (2011), as well as a map of road and railroad construction from the Governor of Puerto Rico (1942).
- (2) **Market access:** following Harris (1954) and many recent studies (e.g., Jedwab et al., 2015), market access is defined as  $\sum_{j \neq i} \frac{N_{j,t}}{d_{ij,t}^\theta}$ , where  $N_{j,t}$  is the population of municipality  $j$  in year  $t$ ,  $d_{ij,t}$  is the travel time between municipalities  $i$  and  $j$ , and  $\theta$  is the distance decay parameter. In my preferred specification,  $\theta = 1$ , but I consider the sensitivity of the results to other values proposed in the literature. Travel time is estimated using historical transport networks modeled in ArcGIS, and speeds for the various forms of communication have been set using various historical sources.
- (3) **Travel time (maximum):** the maximum of the travel times between municipality  $m$  and all other municipalities in 1899. This measure is used as an instrumental variable for

---

<sup>6</sup> Available at <http://www.raremaps.com>

<sup>7</sup> Available at <https://antiquemapsandglobes.com>

<sup>8</sup> Available at <http://www.loc.gov/item/00552205/>

<sup>9</sup> Available at <https://antiquemapsandglobes.com>

market access in 1899. Travel time is estimated by assigning speeds to all types of roads and finding the fastest route between the two cities.

- (4) **Travel time (mean)**: the mean of the travel times between municipality  $m$  and all other municipalities outside of the department in 1899. This measure is used as an instrumental variable for market access in 1899. Travel time is estimated by assigning speeds to all types of roads and finding the fastest route between the two cities.

#### B.1.4 Geographic Controls

- (1) **Crop dummies**: a series of three binary variables for coffee, sugar cane, and tobacco, respectively, equal to 1 if a municipality is in the crop region, as defined by Ayala and Bernabe (2007), and 0 otherwise.
- (2) **Land suitability**: agricultural suitability for rain-fed, low input coffee, sugar cane, and tobacco, respectively, from the Global Agro-Ecological Zones database from the Food and Agricultural Organization of the United Nations. The index is based on average climate of the baseline period 1961-1990 and quantifies to what extent soil conditions match crop requirements. Higher numbers corresponding to better suitability.<sup>10</sup>
- (3) **Mean annual rainfall (inches)**: averaged over 1899 to 1928, from Roberts (1941). Bobonis and Morrow (2014) assigned the simple average of adjacent municipalities' weather measures for municipalities with no available weather stations.
- (4) **Mean altitude (meters)**: generated in ArcGIS by averaging over area using data from the US Geological Survey's National Geospatial Program.<sup>11</sup>
- (5) **Mean land gradient (degrees)**: generated in ArcGIS by averaging over area using data from the Puerto Rico Planning Board. Data come from Bobonis and Morrow (2014).

<sup>10</sup> Available at <http://gaez.fao.org/Main.html>. Last accessed 2/4/2019.

<sup>11</sup> Available at <http://www.usgs.gov>. Last accessed 2/4/2019.

- (6) **Percentage of land cultivated, 1910:** number of acres devoted to the cultivation of coffee, sugar cane, and tobacco, respectively as reported in the 1910 Agricultural Census, divided by total area of the municipality.

## B.2 Additional Tables

Table B.1: Persistent Effects of Historical Market Access on Economic Outcomes: 1910 Agricultural Controls

<i>Dependent variables:</i>	Pop. Density, 1899	Pop. Density, 2010	Urban Extent, 2010	Light Density, 2012
	(1)	(2)	(3)	(4)
Market access, 1899	-0.22 (0.23)	1.34*** (0.22)	1.98*** (0.33)	0.97*** (0.14)
Pop. density, 1899		0.74*** (0.10)	0.65*** (0.14)	0.34*** (0.07)
R-squared	0.21	0.83	0.75	0.76

<i>Dependent variables:</i>	Income, 2012-16	Factories, 1966	Corporations, 2017	Socioeconomic Index, 2013
	(5)	(6)	(7)	(8)
Market access, 1899	0.49*** (0.13)	1.56** (0.69)	2.30*** (0.49)	0.31*** (0.11)
Pop. density, 1899	0.15*** (0.04)	0.74** (0.30)	1.14*** (0.25)	0.04 (0.04)
R-squared	0.65	0.46	0.63	0.34

Notes: All regressions include 76 observations (municipalities). All regressions include the following controls: elevation, distance to the coast, and the percentage of land dedicated to the cultivation of coffee, sugar cane, and tobacco, respectively, in 1910. All variables are expressed as natural logs. Robust standard errors are shown in parentheses. Significance levels are denoted by \* p <.10, \*\* p <.05, \*\*\* p <.01.

Table B.2: Persistent Effects of Historical Market Access on Economic Outcomes: FAO Land Suitability.

<i>Dependent variables:</i>	Pop. Density, 1899	Pop. Density, 2010	Urban Extent, 2010	Light Density, 2012
	(1)	(2)	(3)	(4)
Market access, 1899	-0.07 (0.26)	0.79*** (0.26)	1.34*** (0.35)	0.61*** (0.19)
Pop. density, 1899		0.51*** (0.14)	0.38** (0.18)	0.17 (0.10)
R-squared	0.25	0.84	0.74	0.75

<i>Dependent variables:</i>	Income, 2012-16	Factories, 1966	Corporations, 2017	Socioeconomic Index, 2013
	(5)	(6)	(7)	(8)
Market access, 1899	0.27* (0.14)	1.47* (0.77)	0.89 (0.56)	0.16 (0.11)
Pop. density, 1899	0.07 (0.06)	0.69* (0.35)	0.87*** (0.32)	-0.02 (0.06)
R-squared	0.62	0.43	0.62	0.36

Notes: All regressions include 76 observations (municipalities). All regressions include the following controls: elevation, distance to the coast, and FAO land suitability variables. All variables are expressed as natural logs. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table B.3: Persistent Effects of Historical Market Access on Economic Outcomes: 1899 Development Controls

<i>Dependent variables:</i>	Pop. Density, 1899	Pop. Density, 2010	Urban Extent, 2010	Light Density, 2012
	(1)	(2)	(3)	(4)
Market access, 1899	-0.22 (0.23)	1.34*** (0.22)	1.98*** (0.33)	0.97*** (0.14)
Pop. Density, 1899		0.74*** (0.10)	0.65*** (0.14)	0.34*** (0.07)
R-squared	0.21	0.83	0.75	0.76

<i>Dependent variables:</i>	Income, 2012-16	Factories, 1966	Corporations, 2017	Socioeconomic Index, 2013
	(5)	(6)	(7)	(8)
Market access, 1899	0.49*** (0.13)	1.56** (0.69)	2.30*** (0.49)	0.31*** (0.11)
Pop. Density, 1899	0.15*** (0.04)	0.74** (0.30)	1.14*** (0.25)	0.04 (0.04)
R-squared	0.65	0.46	0.63	0.34

Notes: All regressions include 76 observations (municipalities). All regressions include the following controls: elevation, distance to San Juan, distance to the nearest port, distance to the coast, crop dummies, literacy rate, and the percentage of labor force employed in agriculture, both from 1899. All variables are expressed as natural logs. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table B.4: Persistent Effects of Historical Market Access on Economic Outcomes: Different Values of  $\theta$ 

<i>Dependent variables:</i>	Pop. Density, 2010			Per Capita Income, 2012-16		
	(1)	(2)	(3)	(4)	(5)	(6)
Market access, 1899	0.26*** (0.05)	0.07*** (0.02)	0.04*** (0.01)	0.09*** (0.03)	0.02* (0.01)	0.01 (0.01)
R-squared	0.79	0.78	0.78	0.59	0.56	0.56
$\theta =$	3	8	12	3	8	12

Notes: All regressions include 76 observations (municipalities). All regressions include the following controls: elevation, distance to San Juan, distance to the nearest port, distance to the coast, and crop dummies. All variables are expressed as natural logs. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table B.5: Persistent Effects of Historical Market Access on Economic Outcomes  
—2SLS Results (Maximum Travel Time)

<i>Dependent variables:</i>	Pop. Density, 1899	Pop. Density, 2010	Urban Extent, 2010	Light Density, 2012
	(1)	(2)	(3)	(4)
Market access, 1899	-0.20 (0.36)	0.96*** (0.34)	1.86*** (0.49)	0.69*** (0.25)
Pop. density, 1899		0.64*** (0.10)	0.52*** (0.15)	0.26*** (0.08)
F-statistic (first stage)	72.0	70.2	70.2	70.2

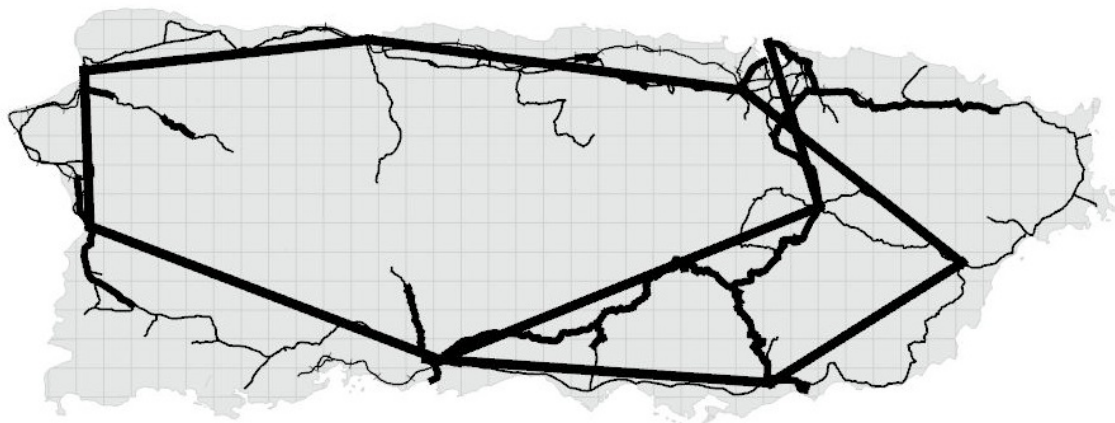
  

<i>Dependent variables:</i>	Income, 2012-16	Factories, 1966	Corporations, 2017	Socioeconomic Index, 2013
	(5)	(6)	(7)	(8)
Market access, 1899	0.11 (0.15)	2.04* (1.16)	1.50* (0.86)	-0.09 (0.14)
Pop. density, 1899	0.09* (0.05)	0.62** (0.26)	0.78*** (0.25)	0.00 (0.05)
F-statistic (first stage)	70.2	70.2	70.2	70.2

Notes: The maximum travel time between a municipality and all other municipalities in 1899 is used as an instrument for market access in 1899. All regressions include 76 observations (municipalities). All regressions include the following controls: distance to San Juan, distance to the nearest port, population density in 1899, elevation, distance to the coast, and FAO crop suitability for coffee, sugar cane, and tobacco. All variables are expressed as natural logs. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

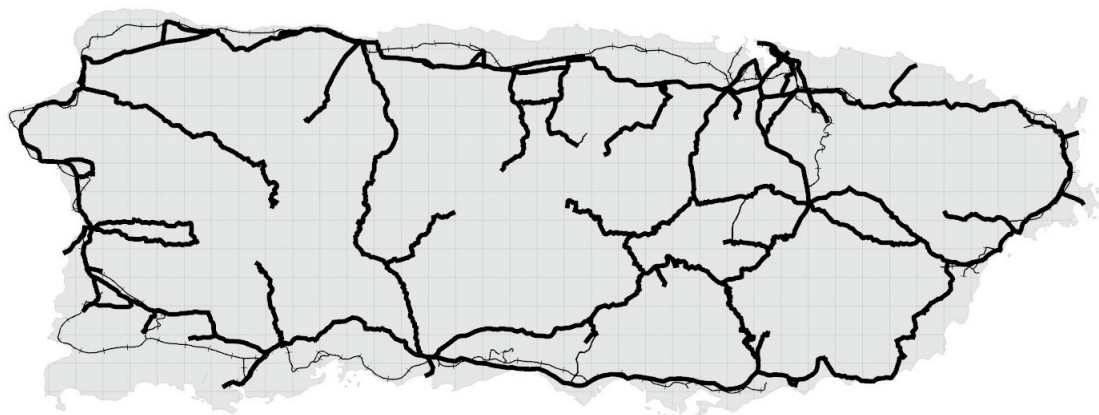
### B.3 Additional Figures

Figure B.1: Transport Network in 1899 and Straight Line Instruments



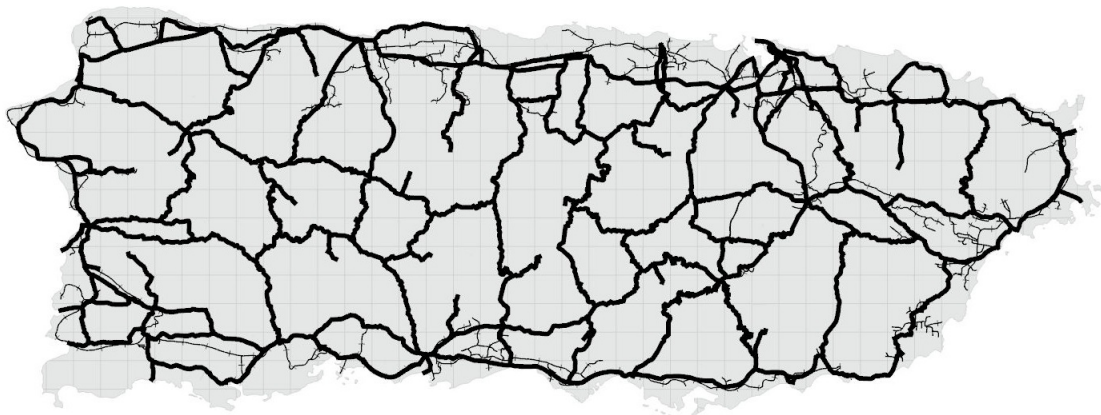
Notes: The transport network is based on maps published by George F. Cram and the US War Department for 1899. The thick black lines are highways; the thin black lines are country roads; and the grey tracked lines are railways. The thickest black lines are straight lines connecting the chief towns of the seven departments and San Juan to Ponce via Caguas, used to construct the instrumental variable. The grey squares are grid cells used as the unit of analysis.

Figure B.2: Transport Network in 1920



Notes: The transport network is based on maps published by Rand McNally for 1920. The thick black lines are highways; the thin black lines are cart roads; and the grey tracked lines are railways. The grey squares are grid cells used as the unit of analysis.

Figure B.3: Transport Network in 1940



Notes: The transport network is based on maps published by Puerto Rico Bureau of Public Works for 1940. The thick black lines are highways; the thin black lines are cart roads; and the grey, tracked lines are railways. The grey squares are grid cells used as the unit of analysis.

#### B.4 Spatial Autocorrelation

Kelly (2019) argues that unusually high  $t$ -statistics in historical persistence research are related to severe spatial autocorrelation in residuals.<sup>12</sup> That is, the results of a large literature drawing connections between modern outcomes and characteristics of the same places in the distance past might be the outcome of fitting spatial noise. The study suggests treating the results of persistence studies, and of spatial regressions more generally, with caution.

I follow Kelly (2019) in reporting Moran's  $I$  statistic, a measure of spatial autocorrelation, for regression residuals. I use spatial contiguity weights such that all areas  $j$  bordering  $i$  are assigned a weight of one, all second-order neighbors (neighbors of neighbors) are assigned a weight of one-half, all other areas are assigned a weight of zero.<sup>13</sup>

Equation 3.3 (excluding observations within 10 km. of network nodes) generates the following z-scores from the Moran test for the corresponding dependent variable: population density in 1920, 7.9; population density in 1940, 5.3; population density in 2010, 6.4; nighttime light density in 2012, 7.5; and urban extent in 2010, 15.2. In other words, for all outcomes we can reject at the

<sup>12</sup> I demonstrate in Section 3.5.1 that my results are robust to correcting for spatial autocorrelation using Conley standard errors.

<sup>13</sup> The conclusions drawn from this section are insensitive to the choice of weights.

1-percent level that residuals are not spatially autocorrelated.

Despite spatial autocorrelation in residuals, the main results do not appear to be spurious. I generate spatial noise using the `geostatsp` R package of Brown (2015) and parameters proposed by Kelly (2019). I set shape  $\kappa = 1$  and correlation range equal to .4 degrees, and half of this north-south, since Puerto Rico is approximately 1.6 degrees east-west. I then run 1,000 simulations in which I replace the  $DistanceToRoads_{1899,i}$  variable with spatial noise. Consistent with Kelly, spatial noise predicts economic outcomes at a statistically significant level at high rates. For instance, spatial noise is statistically significant at the 5-percent level at the following rates: 33.9% of regressions with population density in 1920 as the dependent variable; 28.2%, population density in 1940; 26.7%, population density in 2010; 42.2%, nighttime light density in 2012; 30.4%, urban extent in 2010. However, the placebo outperforms (i.e.  $t$ -statistic is more negative) the true measure of access to roads at the following rates: 3.4% with population density in 1920 as the dependent variable; 0.8%, population density in 1940; 1.0%, population density in 2010; 0.8%, nighttime light density in 2012; and 0.00%, urban extent in 2010.<sup>14</sup> Hence, spatial noise frequently predicts modern outcomes, but it is consistently outperformed by the main explanatory variable.

---

<sup>14</sup> When discussing the performance of spatial noise compared to the true explanatory variable, Kelly (2019) focuses on statistical significance and ignores the sign of the coefficient, which in the case of the effect of roads is theoretically grounded and empirically consistent.

## Appendix C

### Supplemental Material for Chapter 4

#### C.1 Data

##### C.1.1 Dependent Variables

- (1) **Population growth rate:** the average annual growth rate of population from the listed year to the next census, expressed as a natural log. For instance, the growth rate for 1910 is the annual population growth rate from 1910 to 1920. I transcribed data for most Spanish censuses from Vazquez (1978), for US censuses through 1980 from Vazquez as well, and for US censuses from 1980 to 2010 from <http://www.statoids.com/upr.html>. I downloaded the 1790 and 1800 Spanish censuses from Scarano and Curtis (2011). Their dataset includes annual *padrones*, or population counts, from 1779 to 1802, but I use only two years of data in order to keep somewhat consistent spacing between censuses. According to Curtis and Scarano (2011), these are as dependable as any of the population summaries that exist for 18th century Spanish America.
- (2) **Net migration rate:** population in  $t + 1$  minus population in  $t$  minus births plus deaths, all divided by average population over that period. The estimations come from Vázquez Calzada (1978) and are expressed in log form.

### C.1.2 Development Variables

- (1) **Percentage literate, 1899\***: the number of residents ages 10 and up who can read divided by the total population over the age of 10.
- (2) **Percentage literate,  $t^*$** : the number of residents ages 10 and up (or 18 and up after 1980) who can read divided by the total population over the age of 10 (or 18) in year  $t$ . Data are available from 1887 to 2000.
- (3) **Percentage of labor in agriculture, 1899\***: the number of workers, male or female, employed in agriculture, divided by the number of workers in the labor force, as reported in the 1899 Census.
- (4) **Percentage of labor in manufacturing, 1899\***: the number of workers, male or female, employed in manufacturing, divided by the number of workers in the labor force, as reported by the 1899 Census.
- (5) **Index of Municipal Socioeconomic Development, 2013**: produced by the Puerto Rico Planning Board based on per capita income, median family income, families below the poverty level, labor force unemployment rate, population education level, and illiteracy. The index is modeled on the United Nations Human Development Index. For more information, see <https://data.pr.gov/>.

### C.1.3 Geographic Variables

- (1) **Percentage of land cultivated, 1910\***: number of acres devoted to the cultivation of coffee, sugar cane, and tobacco, respectively as reported in the 1910 Agricultural Census, divided by total area of the municipality.
- (2) **Crop dummies**: three binary variables for coffee, sugar cane, and tobacco, respectively, equal to 1 if a municipality is in the crop region, as defined by Ayala and Bernabe (2007), and 0 otherwise.

- (3) **Mean annual rainfall (inches)\***: averaged over 1899 to 1928, from Roberts (1941). Bobonis and Morrow (2014) assigned the simple average of adjacent municipalities' weather measures for municipalities with no available weather stations.
- (4) **Mean altitude (meters)\***: generated in ArcGIS by averaging over area using data from the US Geological Survey's National Geospatial Program at <http://www.usgs.gov>.
- (5) **Mean land gradient (degrees)\***: generated in ArcGIS by averaging over area using data from the Puerto Rico Planning Board. Data come from Bobonis and Morrow (2014).

\*In the regressions, these variables have been normalized to have a mean of zero and a standard deviation equal to one.

## C.2 Standardization of Municipality Boundaries

In Section 4.4, I track the evolution of municipal population densities over two and a half centuries by standardizing city definitions across time. To standardize city definitions I employ an aggregation strategy similar to Horan and Hargis (1995), grouping cities by their modern boundaries into larger units on the basis of historical municipal boundary configurations. I rely on municipality histories from the online Encyclopedia of Puerto Rico published by the Puerto Rican Endowment for the Humanities (2018).

To illustrate my methodology, consider for example the municipalities of Manatí, Barceloneta, and Florida. Manatí was founded in 1738. Barceloneta was founded in 1881 from wards of Manatí. Consequently, the population of Barceloneta is only reported separately from the 1887 census forward but is reported, implicitly, as part of Manatí in all previous censuses. In 1971, Florida separated from Barceloneta and became an independent municipality. The population of Florida is only reported separately from 1980 onward but is included, implicitly, in the population of Manatí from 1765 to 1877 and in the population of Barceloneta from 1887 to 1970.

Suppose that I want to analyze municipal population growth starting in 1765. It is not possible to know the populations of Barceloneta and Florida from 1765 to 2010, so I must include

them as part of Manatí in each census to keep a consistent definition from. As seen in Table C.1, from 1765 to 1877, Barceloneta and Florida are already included in population counts for Manatí in the raw data. From 1887 to 1970, I add together the populations of Manatí and Barceloneta as reported in the census. And from 1980 to 2010, I combine the populations of Manatí, Barceloneta, and Florida to form a definition of Manatí that is consistent from 1765 to 2010. This process would be the same for these three cities if I wanted to start my analysis in any year prior the founding of Barceloneta.

Consider instead that I want to analyze city growth from 1887 onward, since this would require less aggregation and thus allow for more observations in each census year. By 1887, Barceloneta had already separated from Manatí. Hence, the population of Manatí requires no change from the population as reported in the census. From 1887 to 1970, the population of Barceloneta is also identical to that reported in the original census files. For 1980 to 2010, I then add the population of Florida — then reported in the census separately rather than as part of Barceloneta — to the population of Barceloneta to keep a consistent definition of Barceloneta. Once again, this process would be the same for these three cities if I wanted to start my analysis in any year prior the founding of Florida. (I never start analysis later than 1899, so Florida is never treated as an independent municipality in this paper).

In cases in which I evaluate population density, I aggregate land area from the US Census as well. All other variables, such as land use by crop type, are aggregated using the same procedure where possible.

To be clear, the entire island is included regardless of the year in which I start the analysis. The only difference, depending on the year in which I start, is the degree of aggregation. It is thus not possible to run regressions using an unbalanced panel of cities. If I start my analysis in, say, 1827 using a level of aggregation consistent with city definitions in 1827, I do not include prior years in the analysis.

To see municipal definitions by level of aggregation, see Tables C.2 and C.3.

Table C.1: Example of Aggregation of City Populations: Barceloneta, Florida, and Manatí

<b>First Year of Analysis:</b>	Raw data			1765	1887	
<u>Year</u>	<u>Barce- loneta</u>	<u>Florida</u>	<u>Manatí</u>	<u>Manatí</u>	<u>Barce- loneta</u>	<u>Manatí</u>
1765	n/a	n/a	2475	2475	-	-
1776	n/a	n/a	3096	3096	-	-
1790	n/a	n/a	5402	5402	-	-
1800	n/a	n/a	5963	5963	-	-
1827	n/a	n/a	5729	5729	-	-
1846	n/a	n/a	11598	11598	-	-
1860	n/a	n/a	11559	11559	-	-
1877	n/a	n/a	16218	16218	-	-
1887	6246	n/a	8191	14437	6246	8191
1897	7835	n/a	12630	20465	7835	12630
1899	9357	n/a	13989	23346	9357	13989
1910	11644	n/a	17240	28884	11644	17240
1920	13442	n/a	20100	33542	13442	20100
1930	15751	n/a	24838	40589	15751	24838
1935	17377	n/a	27674	45051	17377	27674
1940	18545	n/a	29366	47911	18545	29366
1950	19897	n/a	30449	50346	19897	30449
1960	19334	n/a	29354	48688	19334	29354
1970	20792	n/a	30559	51351	20792	30559
1980	18942	7232	36562	62736	26174	36562
1990	20947	8689	38692	68328	29636	38692
2000	22322	12367	45409	80098	34689	45409
2010	24816	12680	44113	81609	37496	44113

This table illustrates how populations are aggregated to standardize municipality definitions over time using the example of Barceloneta, Florida, and Manatí for analysis starting in the years 1765 and 1887, respectively. Numbers appearing below city names are population counts. For a detailed explanation of the aggregation strategy, including a description of this example, see above in Appendix C.2



Table C.3: Municipality Standardization Decisions (cont.)

Municipality	First Year of Analysis (Level of Aggregation)								
	1765	1776	1827	1846	1860	1877	1887	1897	1899-2010
Sabana Grande	San Germán	San Germán							
Salinas	Coamo	Coamo	Coamo	Coamo					
San Germán									
San Juan									
San Lorenzo	Caguas	Caguas							
San Sebastián									
Santa Isabel	Coamo	Coamo	Coamo						
Toa Alta	San Juan	San Juan							
Toa Baja									
Trujillo Alto	San Juan	San Juan							
Utua									
Vega Alta	Manatí								
Vega Baja	Manatí								
Vieques*	n/a	n/a	n/a						
Villalba	Ponce	Ponce	Juana Díaz	Juana Díaz	Juana Díaz	Juana Díaz	Juana Díaz	Juana Díaz	Juana Díaz
Yabucoa	Humacao	Humacao							
Yauco									

Notes: The municipality name listed in each column indicates to which other municipality a population is assigned for analysis beginning in a given census year. Blank spaces mean that that row's population has not been assigned to another municipality for analysis beginning in that census year. For example, if analysis begins in 1765 (or in 1776), the population of Adjuntas is assigned to Utuado (that is, aggregated with Utuado) for each year of that analysis. If analysis begins in any year after 1776, Adjuntas is treated as a distinct entity from Utuado. To continue the example, for analysis beginning in 1765, Utuado is defined as an aggregation of (modern) Adjuntas, Jayuya, and Utuado. For a detailed explanation of the aggregation strategy, see above in Appendix C.2. \*Vieques and Culebra — the only municipalities not connected to the main island — are not included in analysis from 1765 through 1827.

†Río Piedras was consolidated with the municipality of San Juan in 1951 and is considered part of San Juan throughout my analysis.

### C.3 Correlates of Net Migration

An implied assumption in Section 4.4.3 is that differences in population growth rates across municipalities were driven by migration. But if changes in natality and/or mortality differ across municipalities over time, they may be responsible. To test the robustness of the results, in this subsection I run regressions with net migration as the dependent variable. Between 1899 and 1940, about 326 thousand migrants left the countryside for urban areas, and another 703 thousand did so between 1940 and 1970 (Vázquez Calzada, 1978, p. 310). Despite unrestricted access to the US, there was minimal external migration to the mainland before WWII. In 1940, only about 70 thousand Puerto Ricans lived in the 48 states plus DC. By 1970, however, one-fourth of those born in Puerto Rico lived on the mainland (Ruggles et al., 2018). Hence, estimates of net migration

by municipality from 1900 to 1940 mostly capture internal migration, while those for subsequent decades are driven by emigration to the US.

The estimates of net migration come from Vázquez Calzada (1978) and are equal to population in  $t + 1$  minus population in  $t$  minus births plus deaths in that municipality over that period. The only periods estimated are 1899 to 1930, 1930 to 1940, 1940 to 1950, 1950 to 1960, and 1960 to 1970. The natality and mortality data necessary to estimate internal migration are incomplete before 1930, so estimates for individual decades are not available.

The results appear in Table C.4 and corroborate the results from the previous subsection. Net migration is highly negatively correlated with agricultural employment in 1899 from 1899 to 1960, such that agricultural employment is the best predictor of net migration in each column other than the last. This pattern remains essentially unchanged in omitted results in which the dependent variable is replaced with one of the other three development variables. These results show that labor gravitated towards more developed areas from 1899 to 1970.

Consistent with the results from the estimations based on population growth, net migration in the three decades following US annexation was positively correlated with tobacco production and negatively correlated with coffee production. Panel B also suggests that net migration was positively correlated with sugar production. In the 1930s, the tobacco industry collapsed and workers left the region, while faster growth in the sugar region proceeded forward. Throughout the 1950s and 1960s, the growth of the sugar region continued to outpace that of the coffee region, but it is worth remembering that net migration in these decades mostly reflects external migration to the US mainland.

Table C.4: Correlates of Net Migration

Year	1899-1930	1930-40	1940-50	1950-60	1960-70
Panel A: Crop Dummies					
Agricultural Employment, 1899	-0.25*** (0.02)	-0.66*** (0.06)	-0.80*** (0.09)	-0.67*** (0.07)	-0.16 (0.11)
Sugar	0.23*** (0.07)	0.25 (0.22)	0.70** (0.27)	0.97*** (0.30)	1.87*** (0.52)
Tobacco	0.59*** (0.15)	-0.36 (0.28)	0.27 (0.34)	0.91* (0.50)	0.71 (0.75)
R-squared	0.78	0.76	0.76	0.63	0.30
Panel B: Land Use in 1910					
Agricultural Employment, 1899	-0.25*** (0.02)	-0.66*** (0.08)	-0.82*** (0.10)	-0.66*** (0.07)	-0.22* (0.12)
Land in Sugar, 1910	-0.03 (0.03)	0.02 (0.11)	-0.07 (0.17)	-0.29** (0.14)	0.26 (0.23)
Land in Tobacco, 1910	0.10*** (0.03)	-0.16 (0.11)	-0.00 (0.09)	-0.04 (0.16)	0.05 (0.40)
Land in Coffee, 1910	-0.10*** (0.03)	-0.05 (0.10)	-0.23** (0.10)	-0.50*** (0.12)	-0.65** (0.28)
R-squared	0.80	0.75	0.75	0.67	0.28
Panel C: Geographic Variables					
Agricultural Employment, 1899	-0.20*** (0.03)	-0.60*** (0.09)	-0.63*** (0.11)	-0.44*** (0.11)	0.16 (0.20)
Land Gradient	-0.17** (0.07)	-0.10 (0.18)	-0.49* (0.26)	-0.73** (0.27)	-0.75 (0.45)
Elevation	0.10 (0.08)	-0.11 (0.20)	0.04 (0.18)	0.21 (0.20)	-0.42 (0.27)
Rainfall	-0.04 (0.02)	-0.08 (0.09)	-0.13 (0.12)	-0.24 (0.18)	-0.18 (0.28)
R-squared	0.71	0.76	0.79	0.66	0.39

Notes: The dependent variable is the annual net migration rate between census years, expressed as a natural log. Each regression includes 51 observations. Each panel represents a separate series of regressions, conducted by census year. The constant term is omitted from the table. Variables are described in Appendix C.1. Observations are weighted by municipal population at the beginning of the period. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

## C.4 Metropolitan Areas in Central America and the Caribbean in 1950

Below are the component units of each metropolitan area in 1950 in the countries shown in Figure 4.8 as defined by International Urban Research (1959).

### Costa Rica

- San José: cantones of Central, Goicoechea, Montes de Oca, Moravia, and Tibás.

### Cuba

- Camagüey: Camagüey city.
- Havana (La Habana): municipios of Bauta, Guanabacoa, Habana, Marianao, Regla, Santa María del Rosario, and Santiago de las Vegas.
- Santiago de Cuba: Santiago de Cuba municipio.

### Dominican Republic

- Ciudad Trujillo (Santo Domingo): Santo Domingo distrito.

### El Salvador

- San Salvador: San Salvador distrito.

### Guatemala

- Guatemala City: Guatemala municipio.

### Haiti

- Port-Au-Prince: communes of Pétionville and Port-au-Prince.

### Honduras

- Tegucigalpa: Distrito Central municipio.

### Nicaragua

- Managua: Managua municipio.

### Panama

- Panama City: Panamá distrito.

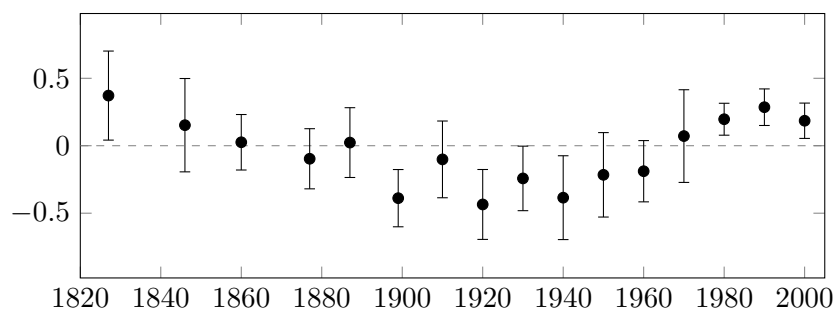
### Puerto Rico

- Ponce: Ponce municipio.
- San Juan: municipios of Bayamón, Cataño, Guaynabo, Río Piedras (annexed to San Juan in 1951), and San Juan.

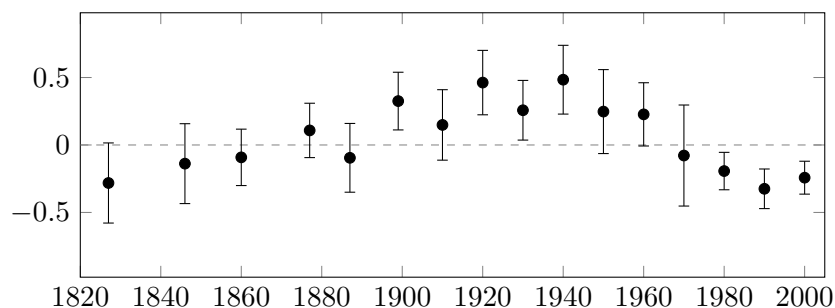
## C.5 Additional Figures

Figure C.1: Development Variables and Population Growth (Unweighted)

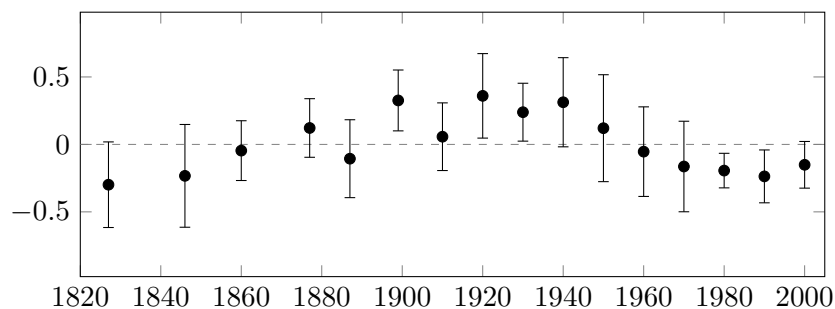
(a) 1899 Agricultural Employment



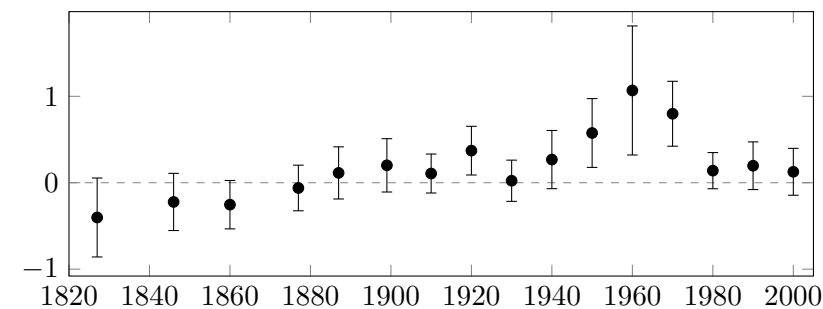
(b) 1899 Literacy Rate



(c) 1899 Manufacturing Employment



(d) 2013 Socioeconomic Development Index



Notes: Regression coefficients and 95 percent confidence intervals come from Equation 2.1, with the population growth rate as the dependent variable. Coefficients correspond to the right-hand side variable shown at the top of each panel. Regressions include 51 observations, based on municipalities aggregated for analysis starting in 1827.

## C.6 Additional Tables

Table C.5: Correlates of Population Growth (Unweighted)

	1827	1846	1860	1877	1887	1899	1910	1920
Agricultural Employment, 1899	0.36** (0.16)	0.16 (0.17)	0.03 (0.10)	-0.10 (0.11)	0.02 (0.13)	-0.39*** (0.10)	-0.10 (0.14)	-0.44*** (0.13)
Sugar	-0.65 (0.53)	-0.48 (0.48)	-0.72 (0.46)	-0.50 (0.30)	-0.76** (0.29)	1.05*** (0.30)	0.27 (0.22)	0.53** (0.26)
Tobacco	-0.42 (0.67)	-1.08 (0.74)	-0.21 (0.63)	0.17 (0.62)	-0.55 (0.34)	1.30*** (0.59)	1.65*** (0.37)	1.56*** (0.47)
Constant	3.19*** (0.47)	2.52*** (0.44)	1.87*** (0.44)	1.25*** (0.26)	1.78*** (0.21)	0.60** (0.26)	1.05*** (0.19)	0.91*** (0.20)
R-squared	0.12	0.05	0.09	0.08	0.09	0.34	0.30	0.30

	1930	1940	1950	1960	1970	1980	1990	2000
Agricultural Employment, 1899	-0.24** (0.12)	-0.38** (0.15)	-0.22 (0.15)	-0.19* (0.11)	0.07 (0.17)	0.20*** (0.06)	0.29*** (0.07)	0.19*** (0.06)
Sugar	0.12 (0.22)	0.64** (0.29)	0.31 (0.30)	1.54*** (0.45)	1.22*** (0.34)	0.49*** (0.17)	0.48** (0.21)	0.53*** (0.18)
Tobacco	-0.31 (0.37)	0.54 (0.39)	0.73** (0.35)	0.87* (0.49)	0.15 (0.34)	0.58** (0.28)	0.03 (0.33)	0.46*** (0.16)
Constant	1.51*** (0.20)	0.62** (0.27)	-0.19 (0.19)	0.00 (0.24)	1.13*** (0.24)	0.81*** (0.13)	0.80*** (0.16)	-0.38*** (0.13)
R-squared	0.22	0.30	0.06	0.14	0.19	0.19	0.20	0.17

Notes: The dependent variable is annual the population growth rate between censuses, expressed as a natural log. Each column represents a separate regression based on Equation 2.1, each with 51 observations. Variables are described in the data appendix. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table C-6: Correlates of Population Growth (Alternative Geographic Variables)

	1827	1846	1860	1877	1887	1899	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
Panel A: Land Use in 1910																
Agr. Emp.	0.22*	-0.05	-0.07	-0.03	-0.12	-0.40***	-0.31***	-0.64***	-0.40***	-0.65***	-0.46***	-0.09	0.42***	0.22***	0.33***	0.22***
	(0.12)	(0.14)	(0.09)	(0.08)	(0.09)	(0.11)	(0.11)	(0.09)	(0.09)	(0.11)	(0.07)	(0.13)	(0.13)	(0.04)	(0.04)	(0.03)
Coffee	-0.05	-0.12	0.06	-0.13	0.10	-0.57***	-0.28***	-0.31**	-0.19*	-0.40***	-0.49***	-0.82***	-0.41*	-0.31***	-0.18*	-0.18**
	(0.18)	(0.23)	(0.15)	(0.11)	(0.12)	(0.16)	(0.10)	(0.13)	(0.10)	(0.11)	(0.12)	(0.26)	(0.21)	(0.09)	(0.10)	(0.08)
Sugar	-0.21	-0.56**	-0.33**	-0.32**	-0.41**	-0.04	-0.48***	-0.28**	-0.22**	-0.33*	-0.48***	-0.10	0.34	-0.15	-0.11	-0.03
	(0.22)	(0.25)	(0.14)	(0.14)	(0.16)	(0.12)	(0.13)	(0.14)	(0.10)	(0.17)	(0.13)	(0.21)	(0.23)	(0.09)	(0.11)	(0.09)
Tobacco	0.01	0.04	0.11	-0.17	0.08	0.14	0.25***	0.28***	-0.30***	-0.14*	-0.10	-0.09	-0.22	-0.03	-0.11	0.07
	(0.16)	(0.21)	(0.11)	(0.22)	(0.12)	(0.09)	(0.08)	(0.09)	(0.09)	(0.08)	(0.13)	(0.41)	(0.21)	(0.07)	(0.07)	(0.06)
Constant	2.42***	1.81***	1.23***	0.92***	1.26***	1.39***	1.38***	1.40***	1.63***	1.12***	0.10	1.17***	2.01***	1.13***	1.03***	-0.06
	(0.19)	(0.29)	(0.17)	(0.14)	(0.15)	(0.10)	(0.08)	(0.10)	(0.07)	(0.11)	(0.14)	(0.30)	(0.21)	(0.09)	(0.08)	(0.08)
R-squared	0.10	0.11	0.14	0.09	0.19	0.62	0.60	0.68	0.70	0.73	0.63	0.23	0.39	0.44	0.49	0.39
Panel B: Geographic Variables																
Agr. Emp.	0.24**	-0.19	-0.18**	-0.13	-0.13	-0.20	-0.23	-0.51***	-0.42***	-0.52***	-0.35***	0.23	0.64***	0.20***	0.34***	0.27***
	(0.11)	(0.20)	(0.09)	(0.09)	(0.11)	(0.15)	(0.16)	(0.11)	(0.10)	(0.13)	(0.12)	(0.24)	(0.16)	(0.04)	(0.06)	(0.05)
Gradient	-0.72***	0.31	0.18	-0.12	0.28	-0.71**	-0.57**	-0.63**	-0.21	-0.71***	-0.68**	-1.04*	-0.17	-0.22*	0.02	-0.20
	(0.24)	(0.40)	(0.23)	(0.18)	(0.21)	(0.29)	(0.24)	(0.24)	(0.18)	(0.24)	(0.29)	(0.55)	(0.37)	(0.12)	(0.17)	(0.14)
Altitude	1.10***	0.29	0.42	0.56***	0.12	0.20	0.59**	0.44*	0.06	0.34*	0.44*	0.06	-0.76***	0.07	-0.25	0.10
	(0.36)	(0.38)	(0.27)	(0.19)	(0.19)	(0.33)	(0.24)	(0.22)	(0.19)	(0.18)	(0.24)	(0.27)	(0.26)	(0.14)	(0.16)	(0.14)
Rain	-0.04	0.17	0.04	-0.02	-0.12	-0.11	0.10	0.08	0.08	-0.02	-0.17	-0.23	-0.26	0.25***	0.03	0.00
	(0.14)	(0.24)	(0.11)	(0.13)	(0.20)	(0.14)	(0.12)	(0.10)	(0.10)	(0.12)	(0.14)	(0.29)	(0.24)	(0.08)	(0.06)	(0.07)
Constant	2.44***	1.76***	1.21***	0.85***	1.26***	1.30***	1.35***	1.39***	1.62***	1.10***	0.05	1.10***	1.95***	1.14***	1.03***	-0.08
	(0.17)	(0.32)	(0.18)	(0.13)	(0.15)	(0.12)	(0.11)	(0.14)	(0.07)	(0.11)	(0.14)	(0.27)	(0.16)	(0.08)	(0.08)	(0.09)
R-squared	0.26	0.09	0.25	0.23	0.15	0.46	0.31	0.59	0.65	0.75	0.55	0.27	0.52	0.51	0.51	0.34

Notes: Each regression includes 51 observations. Observations are weighted by municipal population at the beginning of the period. Each panel represents a separate series of regressions, conducted by census year. Variables are described in Appendix C.1. Robust standard errors are shown in parentheses. Significance levels are denoted by \* p < .10, \*\* p < .05, \*\*\* p < .01.

## Appendix D

### Supplemental Material for Chapter 5

#### D.1 Additional Tables

Table D.1: Changes in Height Across Birth Cohorts in the PRHP and PREHCO (Quantile Regressions)

	(1)	(2)	(3)	(4)	(5)	(6)
Born 1896-00	2.54** (1.05)	2.54*** (0.47)	0.00 (1.15)			
Born 1901-05	0.00 (1.00)	0.00 (0.45)	2.54** (1.11)			
Born 1906-10	2.54** (1.00)	2.54*** (0.45)	2.54** (1.10)			
Born 1911-15	2.54** (1.00)	2.54*** (0.45)	2.54** (1.10)			
Born 1916-20	2.54** (1.00)	2.54*** (0.45)	2.54** (1.10)	0.00 (3.05)	0.00 (1.60)	5.08*** (1.53)
Born 1921-25	2.54** (1.11)	2.54*** (0.50)	2.54** (1.22)	0.00 (3.01)	2.54* (1.49)	5.08*** (1.72)
Born 1926-30	0.00 (1.58)	2.54*** (0.71)	5.08*** (1.74)	3.18 (2.97)	2.54** (1.22)	6.35*** (1.65)
Born 1931-35				3.18 (2.93)	5.08*** (1.16)	7.62*** (1.58)
Born 1936-40				5.08* (2.91)	5.08*** (1.25)	7.62*** (1.46)
Born 1941-42				5.08* (3.05)	5.08*** (1.32)	7.62*** (1.65)
Constant	157.48*** (0.98)	162.56*** (0.44)	167.64*** (1.08)	157.48*** (2.85)	162.56*** (1.05)	165.10*** (1.42)
Quantile	0.20	0.50	0.80	0.20	0.50	0.80
Observations	9,805	9,805	9,805	1,905	1,905	1,905

Notes: Results are derived from quantile regressions. The dependent variable is standing height in centimeters. Data include men born between 1886 and 1930 for the PRHP and between 1896 and 1942 in the PREHCO. The data were collected in 1965 for the PRHP and in 2002-03 for the PREHCO. Robust standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table D.2: Changes in Predicted Height Across Birth Cohorts in the 1970 US and Puerto Rican Censuses

	(1)	(2)	(3)
Born 1896-00	0.60 (0.79)	0.62 (0.77)	0.06 (0.83)
Born 1901-05	0.36 (0.76)	0.36 (0.73)	-0.08 (0.80)
Born 1906-10	1.11 (0.75)	1.14 (0.73)	0.44 (0.79)
Born 1911-15	1.62** (0.74)	1.69** (0.73)	0.80 (0.78)
Born 1916-20	1.81** (0.74)	1.83** (0.73)	1.09 (0.78)
Born 1921-25	2.67*** (0.82)	2.80*** (0.80)	1.70** (0.86)
Born 1926-30	2.27* (1.31)	2.40* (1.28)	1.30 (1.31)
Constant	163.43*** (0.74)	163.35*** (0.72)	164.32*** (0.77)
Sample	US & PR	PR	US
Observations	4,235	3,122	1,113
R-squared	0.36	0.39	0.22

Notes: The dependent variable is predicted height in centimeters estimated using coefficients from a regression of height on birth cohort and educational attainment in the PRHHP. Data include men born between 1886 and 1930 in Puerto Rico who appear in the one-percent samples of the US and Puerto Rican censuses from Ruggles et al. (2018). The US sample includes migrants, and the Puerto Rican sample includes non-migrants. Bootstrapped standard errors are shown in parentheses. Significance levels are denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

Table D.3: Average (Imputed) Years of Education of Puerto Rican Men by Birth Year

Birth Years	(A) 1970 Census		(B) PRHHP		(C) PREHCO	
	Mean	N	Mean	N	Mean	N
Born 1886-00	3.4	482	3.7	679	2.5	1
Born 1901-05	4.2	374	4.5	1,742	4.1	5
Born 1906-10	5.1	423	5.3	2,113	6.2	20
Born 1911-15	5.4	475	6.5	2,676	5.1	74
Born 1916-20	6.9	544	7.5	2,245	6.0	164
Born 1921-25	7.8	536	6.8	297	6.8	241
Born 1926-30	8.3	612	7.8	51	8.4	329
Born 1931-35	9.0	731	-	-	9.1	408
Born 1936-40	9.2	734	-	-	9.0	474
Born 1941-42	9.5	312	-	-	9.4	153

Notes: The PRHHP does not include detailed information about educational attainment. For comparability, in all three data sets years of education are imputed as follows: no schooling (0 years), primary (2.5), middle (6.5), attended high school (10), graduated high school (12), attended college (14), and graduated college or more (16).