

UNIVERSITY OF OTTAWA

DOCTORAL THESIS

Essays in Socioeconomic Health
Inequalities

Author:

Khadija BCHI

Supervisors:

Myra YAZBECK

Paul MAKDISSI

*Thesis submitted to the University of Ottawa in partial Fulfillment of the
requirements for the degree of Doctor of Philosophy in Economics*

Department of Economics

Faculty of Social Sciences

University of Ottawa

Abstract

In the first chapter, we derive the expressions for the recentered influence function (RIF) of the absolute and relative concentration curves. We also adapt the reweighting approach of DiNardo, Fortin, and Lemieux (1996) to construct counterfactual concentration curves. We then apply these methods to decompose changes in these curves between 2000 and 2020. This framework provides a comprehensive view of changes in health inequality along the socioeconomic distribution, unlike the decomposition of a synthetic index. To illustrate the approach, we use U.S. data on body mass index (BMI) and cigarette consumption from the National Health Interview Survey (NHIS).

The second chapter examines the impact of informal (de facto) versus legalized (de jure) segregation under Jim Crow laws on the cognitive health of older Black Americans born before 1964. We combine a Difference-in-Differences (DiD) framework with RIF regression applied to health concentration curves. This approach goes beyond average effects and captures distributional impacts across socioeconomic ranks. Using data from the 2016 wave of the Health and Retirement Study, we find that cognitive health benefits of being born in regions with only de facto segregation extend up to the 85th income percentile. No significant effects are found among the highest-income individuals, resulting in an insignificant average treatment effect. These findings highlight the importance of a distributional approach beyond average effects.

The third chapter examines malnutrition, a major public health issue with significant economic consequences in Egypt. Focusing on 2000–2021, we analyze socioeconomic disparities in malnutrition among Egyptian women of reproductive age. We use the health concentration index and Wagstaff’s health shortfall index to study the evolution of these disparities. We also investigate key determinants of these inequalities using a RIF regression approach. Our findings provide a comprehensive perspective on nutritional inequality patterns and their drivers, contributing to evidence-based public health policy discussions.

Declaration of Authorship

I, Khadija BCHI, declare that this thesis titled, “Essays in Socioeconomic Health Inequalities” and the work presented in it are my own. The first and second chapters of this thesis were completed jointly with Myra Yazbeck and Paul Makdissi, with my contribution being equal to theirs. The third chapter was conducted in collaboration with Myra Yazbeck, Paul Makdissi, and Mohamad A. Khaled, with my contribution being equal to theirs as well. I hereby acknowledge the contributions of my thesis supervisors, Myra Yazbeck and Paul Makdissi, to the research related to all three chapters. I also acknowledge the contribution of Mohamad A. Khaled to the research related to the third chapter. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the University of Ottawa.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at the University of Ottawa or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Khadija Bchi

Date: 2026-04-14

Acknowledgements

I would like to begin by expressing my profound gratitude to my supervisors, Myra Yazbeck and Paul Makdissi, whose expertise, guidance, and steadfast patience have been invaluable throughout this journey. Their encouragement, insightful feedback, and mentorship have not only shaped my research but also contributed significantly to my personal and professional growth. I am deeply appreciative of the time and dedication they devoted to supporting me in completing this important chapter of my life. I am equally grateful for their constant support, sense of humor, and the memorable anecdotes we shared along the way.

I extend my sincere thanks to my thesis committee members, Myra Mohnen, Louis-Philippe Béland, and Roland Pongou, for their thoughtful suggestions, ideas, and constructive feedback. I am also grateful to the external examiner, Catherine Haeck, for agreeing to evaluate my thesis and for her valuable comments and recommendations. I also wish to thank Louis-Philippe Morin, Director of the PhD program, for the insightful comments he shared during workshops, for his patience, and for his support throughout my doctoral studies.

I am deeply thankful to the members of the Economics Department, including both academic and administrative staff, for their assistance, patience, and constant support throughout this process.

Finally, I cannot conclude without expressing my heartfelt gratitude to my family—my husband, my children, my mother, my sisters, and my brothers—for their unconditional love, support, patience, and understanding during this journey. This accomplishment would not have been possible without them.

*In memory of my father, who never had the chance to see me
pursue a Ph.D.;*

*To Rhita, Ismail, and Abderazzak, who endured both the
challenges and the joys of this journey;*

*To my mother, brothers, sisters, nephews, and nieces, whose first
question was always, “Have you finished?”— and yes, I did;
To Sara, my sister and my friend, who always stood by my side.*

Contents

Abstract	ii
Acknowledgements	iv
List of Figures	ix
List of Tables	xiii
General Introduction	1
1 Decomposing Changes in Socioeconomic Health Inequality	4
1.1 Introduction	4
1.2 Theoretical framework	6
1.2.1 Definition of the absolute and the relative concentration curves .	6
1.2.2 Decomposition terms	8
1.2.3 Reweighting approach	11
1.2.4 Recentered influence function	13
1.3 Empirical illustration	16
1.3.1 Data and descriptive statistics	16
1.3.2 Distributional decomposition of changes in cigarette consumption	18
Reweighted relative concentration curves	19
Reweighted absolute concentration curves	20
RIF regression for the relative concentration curves	22
RIF regression for the absolute concentration curves	23
1.3.3 Distributional decomposition of change in overweightness	25
Reweighted absolute and relative concentration curves	25

	RIF regression for the absolute and the relative concentration	
	curves	26
1.4	Conclusion	28
	Appendix	30
2	Cognitive Health: De Facto vs De Jure Segregation	49
2.1	Introduction	49
2.2	Background and literature review	51
2.2.1	Jim Crow laws in the South	51
2.2.2	Literature review	52
2.3	Theoretical framework	54
2.3.1	Socioeconomic health inequality measures	55
2.3.2	Identification	57
	RIF approach of the concentration curves	57
	Causal distributional treatment effect model	59
2.4	Data and descriptive statistics	59
2.4.1	Health outcome variables	60
2.4.2	Sample restriction and descriptive statistics	62
2.5	Empirical strategy	63
2.5.1	Econometric models	63
2.5.2	Reweighting approach	66
2.6	Results	67
2.6.1	Descriptive statistics of the outcome variables	67
2.6.2	Estimated treatment effect on health achievement	68
2.6.3	Estimated treatment effect on socioeconomic health inequality	71
2.7	Robustness check	72
2.7.1	Age	72
2.7.2	Migration	72
2.7.3	Placebo test	73
2.8	Conclusion	73
	Appendix	77
3	The dynamics of socioeconomic inequality of malnutrition in Egypt	101

3.1	Introduction	101
3.2	Methodology	102
3.3	Socioeconomic inequalities in anemia and overweightness in Egypt . . .	105
3.3.1	The Egyptian context	105
3.3.2	The data	106
3.3.3	Malnutrition shortfall	109
3.3.4	Socioeconomic inequality in malnutrition	112
3.4	Conclusion	114
	Appendix	116
	General Conclusion	130
	Bibliography	132

List of Figures

1.1	Relative concentration curves	35
1.2	Absolute concentration curves	36
1.3	Counterfactual relative concentration curves- Cigarette consumption (Model 2)	36
1.4	Counterfactual relative concentration curves-Cigarette consumption (Two models)	37
1.5	Endowment and structural effects curves-Relative concentration curves (RCC) of cigarette consumption	37
1.6	Counterfactual absolute concentration curves-Cigarette consumption (Model 2)	38
1.7	Counterfactual absolute concentration curves-Cigarette consumption (Two models)	38
1.8	Endowment and structural effects curves-Absolute concentration curves (ACC) of cigarette consumption	39
1.9	RIF model fit of the relative concentration curves-Cigarette consumption	39
1.10	Structural and endowment effects-Relative concentration curves of cigarette consumption	40
1.11	Contribution of each variable to the structural and the endowment effects in Model 2-Relative concentration curve of cigarette consumption	41
1.12	RIF model fit of the absolute concentration curves-Cigarette consumption	42
1.13	Structural and endowment effects-Absolute concentration curves of cigarette consumption	42
1.14	Contribution of each variable to the structural and the endowment effects in Model 2-Absolute concentration curve of cigarette consumption	43
1.15	Counterfactual curves-Overweightness (Model 2)	44

1.16	Endowment and structural effects curves-Overweightness	45
1.17	RIF model fit of the relative and the absolute concentration curves- Overweightness	45
1.18	Endowment and structural effects curves	46
1.19	Contribution of each variable to the structural and the endowment effects in Model 2-Relative concentration curve of overweightness	47
1.20	Contribution of each variable to the structural and the endowment effects in Model 2-Absolute concentration curve of overweightness	48
2.1	Map of Jim Crow America	80
2.2	Relative concentration curve and concentration index	81
2.3	Absolute concentration curve	81
2.4	Total cognition summary score and its components	82
2.5	Absolute and relative concentration curve-Cognition score	82
2.6	Absolute and relative concentration curve-Mental status score	83
2.7	Absolute and relative concentration curve-Word recall score	83
2.8	Absolute and relative concentration curve-Number series score	84
2.9	Treatment effect on the treated on the absolute concentration curve of Cognition summary score	84
2.10	Treatment effect on the treated for the absolute concentration curve of Cognition summary score	85
2.11	Treatment effect on the treated for the absolute concentration curve of mental status score	85
2.12	Treatment effect on the treated for the absolute concentration curve of word recall score	86
2.13	Treatment effect on the treated for the absolute concentration curve of number series score	86
2.14	Treatment effect on the treated for the relative concentration curve of cognition score	87
2.15	Treatment effect on the treated for the relative concentration curve of mental status score	87

2.16	Treatment effect on the treated for the relative concentration curve of word recall score	88
2.17	Treatment effect on the treated for the relative concentration curve of number series score	88
2.18	Treatment effect on the treated for the absolute concentration curve-Age 60-80 years	89
2.19	Treatment effect on the treated for the relative concentration curve-Age 60-80 years	90
2.20	Treatment effect on the treated for the absolute concentration curve-No migration (the same region JC all waves)	91
2.21	Treatment effect on the treated for the relative concentration curve-No migration (the same region JC all waves)	92
2.22	Treatment effect on the treated for the absolute concentration curve-Placebo in the control group	93
2.23	Treatment effect on the treated for the relative concentration curve-Placebo in the control group	94
2.24	Treatment effect on the treated for the absolute concentration curve-Placebo in the treatment group	95
2.25	Treatment effect on the treated for the relative concentration curve-Placebo in the treatment group	96
2.26	Treatment effect on the treated for the absolute concentration curve-Non-reweighted	97
2.27	Treatment effect on the treated for the relative concentration curve-Non reweighted	98
2.28	Treatment effect on the untreated for the absolute concentration curve .	99
2.29	Treatment effect on the untreated for the relative concentration curve .	100
3.1	Proportion of women with malnutrition deficiencies	117
3.2	Shortfall in anemia and dual burden of anemia and overweightness . . .	124
3.3	Shortfall in overweightness	125
3.4	Socioeconomic inequalities in malnutrition	126
3.5	Overweightness by age	127

3.6 Anemia by age 128

3.7 Dual burden by age 129

List of Tables

1.1	Descriptive statistics	32
1.2	Dominance tests	33
1.3	Dominance tests for counterfactuals of the relative concentration curves	33
1.4	Dominance tests for counterfactuals of the absolute concentration curves	34
1.5	Dominance tests for counterfactuals of the relative concentration curves of overweightness	34
1.6	Dominance tests for counterfactuals of the absolute concentration curves of overweightness	35
2.1	Weighted descriptive statistics	77
2.2	Model treatment effect on treated of $C_A(q)$ of cognition summary score	78
2.3	Weighted descriptive statistics for individuals aged 60 to 80 years	79
3.1	Descriptive statistics by year	116
3.2	Anemia shortfall index regressions	118
3.3	Overweightness shortfall index regressions	119
3.4	Dual burden shortfall index regressions	120
3.5	Anemia concentration index regressions	121
3.6	Overweightness concentration index regressions	122
3.7	Dual burden concentration index regressions	123

General Introduction

Socioeconomic health inequality remains a critical issue in public health and social policy, as it reflects disparities in health across income distributions and social strata. Understanding both the magnitude of these inequalities and the mechanisms driving them is essential for designing effective interventions. A substantial body of research has focused on their measurement. One common approach involves index-based measures, such as the concentration index (Wagstaff, Paci, and Doorslaer, 1991) and the health achievement index (Wagstaff, 2002), which provide valuable insights into the distribution of health outcomes across socioeconomic ranks. However, the decision of comparing health inequalities between two distributions using these indices may vary depending on the chosen mathematical form of the index.

The second approach is the dominance approach, which provides a robust ranking that is independent of the choice of index (Khaled, Makdissi, and Yazbeck, 2018; Makdissi and Yazbeck, 2014). This method compares the health concentration curves of two distributions, including both relative and absolute concentration curves. The relative concentration curve plots the cumulative proportion of health H against the cumulative proportion of the population ranked by income Y , from the most disadvantaged to the least disadvantaged. The absolute concentration curve incorporates both the average level of health and socioeconomic inequality, representing the cumulative contribution of individuals ranked q or lower to the overall average health.

In addition to comparing health inequalities, assessing policies and evaluating their impact on health disparities are of great importance to researchers and policymakers. Standard regression frameworks, such as OLS (ordinary least squares), typically estimate treatment effects using the mean of health outcomes. However, focusing solely on averages can obscure the heterogeneity of effects across different socioeconomic

ranks. Moreover, these frameworks are not readily applicable when the statistic of interest is a distributional measure other than the mean. To address this limitation, Firpo, Fortin, and Lemieux (2009) proposed the recentered influence function (RIF) approach, which bridges inherently nonlinear distributional statistics with standard regression methods, initially applied to univariate distributional statistics. This approach was later extended to the concentration index—a bivariate functional of the joint distribution of health and income—by Heckley, Gerdtham, and Kjellsson (2016) and Erreygers and Van Ourti (2011), and to the health achievement (or shortfall) index by Abu-Ismaïl et al. (2020).

However, analyses based on aggregate measures, such as commonly used indices, may mask heterogeneous impacts across socioeconomic ranks. The factors shaping health inequalities—and the impact of social and economic policy changes—may vary significantly along the distribution. This underscores the need for distributional approaches that go beyond summary indices to capture variations across socioeconomic groups. Such approaches are particularly valuable for assessing the impact of economic and social policies. By examining how health advantages and disadvantages are distributed, rather than focusing exclusively on averages, these methods allow researchers and policymakers to identify which groups benefit most from policy changes and which remain vulnerable.

The three essays presented in this thesis share a common methodological foundation: the use of distributional measures of socioeconomic health inequality combined with RIF regressions to analyze the determinants of health inequality and health achievement/shortfall. While the empirical questions address different policy issues, they converge on a common objective: understanding the levels and drivers of health inequality and the distribution of health achievement/shortfall.

Specifically, the main contribution of Chapter 1 is methodological. By extending the RIF proposed by Firpo, Fortin, and Lemieux (2009) to absolute and relative concentration curves, it develops a framework for decomposing changes in these curves between two groups or time periods. This framework combines reweighting techniques and RIF-based Oaxaca-Blinder decompositions to distinguish between composition and structural effects, providing a detailed view of the factors driving changes in

health inequality and health achievement across the socioeconomic distribution.

Building on Chapter 1, the Chapter 2 applies the RIF of absolute and relative concentration curves in a difference-in-differences (DID) setting, offering a framework for distributional analysis while still incorporating the average treatment effect. It specifically investigates the long-term effect of de facto segregation versus de jure segregation under Jim Crow laws. Our findings indicate that the cognitive health benefits of being born in a region where only de facto segregation existed were observed across the majority of socioeconomic ranks, up to the 85th income percentile. However, no significant effect was observed at the top of the distribution, resulting in a non-significant average treatment effect. This analysis not only provides evidence of the long-term consequences of legalized segregation but also highlights the value of using distributional approaches in policy evaluation.

Chapter 3 focuses on malnutrition in Egypt, where it remains a pressing public health challenge with significant health and economic implications. In this chapter, the definition of malnutrition extends beyond undernutrition to encompass all forms of nutritional imbalance, including overweight and anemia. This chapter examines trends in overweightness and anemia among women of childbearing age over two decades, highlighting persistent and evolving socioeconomic disparities. Using RIF-regressions of the concentration and shortfall indices, the analysis identifies the determinants of these disparities and explores their implications for women's health.

These chapters demonstrate how distributional analysis can generate richer insights into health inequalities and health achievement/shortfall—insights that are critical for designing policies that not only improve health outcomes but also reduce disparities across socioeconomic groups.

Chapter 1

Decomposing Changes in Socioeconomic Health Inequality

1.1 Introduction

A substantial body of literature on the measurement of socioeconomic health inequality has developed since Wagstaff, Paci, and Doorslaer (1991). Among the most widely used tools are the concentration index and its associated concentration curve, both originally adapted from the income inequality literature. The relative concentration curve relates cumulative health outcomes to socioeconomic status and provides a comprehensive picture of how health is distributed across individuals, from the most disadvantaged to the least. However, concerns arise when socioeconomic health inequalities are particularly pronounced within specific subgroups, prompting researchers and policymakers to seek explanations for disparities between groups. For policymakers, such insights are essential for designing and targeting effective interventions.

One widely used approach to explain differences between groups is decomposition analysis, which aims to identify the factors driving variations in a measure of socioeconomic health inequality across groups or periods ($g = 1, 2$). More specifically, decomposition methods break down the overall change into a component attributable to differences in the distribution of characteristics (the composition effect) and a component attributable to differences in the returns to those characteristics (the structural effect). In this way, decomposition provides insight into which factors are quantitatively important, which are not, and the extent to which each factor contributes to

changes in socioeconomic health inequality. Consequently, this approach offers researchers and policymakers valuable guidance for further investigation into why and how key determinants differ across groups.

The objective of this work is to propose a framework for decomposing changes in the absolute and relative concentration curves between two groups or periods ($g = 1, 2$). Rather than focusing on the decomposition of a synthetic index, as in previous studies (Abu-Ismaïl et al., 2020; Erreygers and Kessels, 2013; Heckley, Gerdtham, and Kjellsson, 2016; Wagstaff, 2002), we adopt a distributional perspective. This framework provides a detailed view of how socioeconomic health inequality evolves along the entire socioeconomic distribution, thereby allowing us to capture potential heterogeneity in the effects of individual characteristics across social ranks.

Implementing such a framework requires constructing a counterfactual distribution, defined as the distribution that would have prevailed for group 2 had individuals' characteristics remained as in group 1. The literature offers a broad range of methods for constructing counterfactuals (Fortin, Lemieux, and Firpo, 2011). One prominent approach is the reweighting method proposed by DiNardo, Fortin, and Lemieux (1996), which relies on a reweighting function to generate a counterfactual density. A second class of methods is based on regression techniques (Chernozhukov, Fernández-Val, and Melly, 2013; Firpo, Fortin, and Lemieux, 2009; Machado and Mata, 2005).

Accordingly, we propose two complementary approaches to decompose the absolute and relative concentration curves. The first adapts the reweighting approach of DiNardo, Fortin, and Lemieux (1996) to construct reweighted concentration curves and perform an aggregate decomposition. The second approach builds on the influence function (IF) and the recentered influence function (RIF) developed by Firpo, Fortin, and Lemieux (2009), combined with the decomposition framework of Oaxaca (1973) and Blinder (1973). This latter approach enables a detailed decomposition, identifying the contribution of each factor to changes in the concentration curves.

This paper contributes to the literature in two main ways. First, it extends the growing body of research on the decomposition of socioeconomic health inequality measures. Second, to the best of our knowledge, it is the first study to propose a framework for

the distributional decomposition of absolute and relative concentration curves using both reweighting and RIF-based approaches, rather than relying on a synthetic index decomposition. Synthetic measures may conceal offsetting effects of individual factors at different points of the distribution—for instance, when a determinant has opposing impacts at the lower and upper ends of the socioeconomic spectrum, resulting in an insignificant average effect. In addition, we derive the expressions for the influence function and the recentered influence function for both concentration curves.

To illustrate the proposed framework, we use data from the National Health Interview Survey (NHIS) in the United States. Our health outcomes are cigarette consumption and overweightness, measured as the deviation of body mass index (BMI) from the *normal* BMI. The analysis includes individual demographic characteristics such as age, race, sex, education, region of residence, and the presence of children in the household.

The remainder of the paper is organized as follows. Section 1.2 presents the theoretical framework and details how the two approaches are applied to decompose the absolute and relative concentration curves. Section 1.3 provides an empirical illustration of cigarette consumption and overweightness in the United States; it is purely descriptive and does not offer policy recommendations. Section 1.4 concludes.

1.2 Theoretical framework

1.2.1 Definition of the absolute and the relative concentration curves

We focus on the absolute and the relative concentration curves as measures of socioeconomic health inequality. The relative concentration curve plots the cumulative proportion of health H against the cumulative proportion of the population ranked by a measure of socioeconomic status Y from the most disadvantaged to the least disadvantaged.

Let H and Y be two random variables that are absolutely continuous with support on the positive half real line ($[0, h_{max}] \times \mathfrak{R}_+$) with density functions f_H and f_Y respectively, and with a joint density function f_{HY} . The concentration curve is viewed as a functional of F_{HY} , the joint cumulative distribution of Y and H . The relative concentration curve is formally defined by:

$$C_R(q, F_{HY}) = \frac{1}{\mu_h} \int_0^{F_Y^{-1}(q)} \int_0^{h_{max}} h f_{Y,H} dh dy, \quad (1.1)$$

where $\mu_h = \int_0^\infty \int_0^{h_{max}} h dF_{H|Y}(h|y) dF_Y(y)$ is the average health, $F_Y(y) = \int_0^y f_Y(u) du$ is the cumulative distribution of Y and $F_Y^{-1}(q) = \inf\{y|q \leq F_Y(y)\}$ is the q -th quantile.

The absolute concentration curve is, however, defined by:

$$C_A(q, F_{HY}) = \int_0^{F_Y^{-1}(q)} \int_0^{h_{max}} h f_{HY} dh dy. \quad (1.2)$$

To estimate the absolute and the relative concentration curves, we follow the framework proposed by Khaled, Makdissi, and Yazbeck (2018). Suppose that a sample of N observations is drawn from $F_{Y,H}$. A natural estimator for the absolute concentration curve is then given by

$$\widehat{C}_A(q, F_{HY}) = \frac{1}{N} \sum_{i=1}^N h_i \mathbb{1}(y_i \leq \widehat{F}_Y^{-1}(q)), \quad (1.3)$$

where

$$\widehat{F}_Y^{-1}(q) = \inf\{y : \widehat{F}_Y(y) \geq q\}, \quad (1.4)$$

and

$$\widehat{F}_Y(y) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(y_i \leq y). \quad (1.5)$$

The estimator of the relative concentration curve is given by:

$$\widehat{C}_R(q, F_{HY}) = \frac{\widehat{C}_A(q, F_{HY})}{\widehat{\mu}_H}, \quad (1.6)$$

where

$$\widehat{\mu}_H = \frac{1}{N} \sum_{i=1}^N h_i. \quad (1.7)$$

Finally, it is possible to construct Hájek estimators to account for survey weights by replacing $1/N$ by $\omega_i / \sum_{j=1}^N \omega_j$, where ω_i denotes the sampling weight of observation i .

1.2.2 Decomposition terms

In this section, we present the decomposition framework. This approach aims to identify the factors driving variations in socioeconomic health inequalities. More specifically, it decomposes the overall change in the distribution of socioeconomic health outcomes between two groups or periods ($g=1, 2$) into a component due to differences in the distribution of characteristics (the composition effect) and a component due to differences in the returns to these characteristics (the structural effect).

The decomposition approach relies on comparing the observed and counterfactual distributions of socioeconomic health outcomes. The counterfactual distribution is the distribution that would have occurred in group 2 if individuals' characteristics had remained as in group 1.

We are interested in decomposing Δ^O , the overall difference in the observed absolute and relative concentration curves between group 2 and group 1. We assume that each group has a joint distribution of health H , income Y , covariates X , and unobservable characteristics U . These variables are assumed to have support $[0, h_{\max}] \times \mathfrak{R}_+ \times \mathcal{X} \times \mathcal{U}$. Let us denote by $f_{H_g Y_g X_g U_g}$ the corresponding joint density and by $F_{H_g Y_g X_g U_g}$ the associated joint cumulative distribution.

For simplicity, the decomposition terms are presented only for the absolute concentration curve in the following section. The corresponding terms for the relative concentration curve can be easily derived from those of the absolute concentration curve. In this context, Equation (1.2) can be rewritten as

$$C_A(q; F_{H_g Y_g X_g U_g}) = \int_{\mathcal{X}} \int_{\mathcal{U}} \int_0^{F_Y^{-1}(q)} \int_0^{h_{\max}} h f_{H_g | Y X U}(h|y, x, u) dh f_{Y_g | X U}(y|x, u) dy f_{U_g | X}(u|x) du dF_{X_g}(x),$$

$$g = 1, 2. \tag{1.8}$$

Note: The use of $dF_{X_g}(x)$ is to allow for discrete covariates.

Hence, the overall change in the observed absolute concentration curves between group 2 and group 1 is:

$$\Delta^O(C_A) = C_A(q; F_{H_2 Y_2 X_2 U_2}) - C_A(q; F_{H_1 Y_1 X_1 U_1}). \tag{1.9}$$

The overall difference can be decomposed into 4 components:

1. the difference in the return to observable characteristics X ;
2. the difference in the return to unobservable characteristics U ;
3. the difference in the distribution of the observable characteristics X ;
4. the difference in the distribution of the unobservable characteristics U ;

However, because it is difficult to disentangle the first two effects, the returns to observable and unobservable characteristics are combined into what is referred to as the structural effect, denoted by $\Delta^S(C_A)$. Under certain assumptions¹, the overall difference can be expressed as:

$$\Delta^O(C_A) = \Delta^S(C_A) + \Delta^X(C_A) + \Delta^U(C_A). \quad (1.10)$$

where $\Delta^X(C_A)$ and $\Delta^U(C_A)$ respectively refer to the contribution of difference in the distributions of X (endowment or composition effect) and U to the overall distributional change $\Delta^O(C_A)$.

Moreover, this type of decomposition requires constructing a counterfactual distribution. When group 2 is taken as the anchor point, the counterfactual distribution corresponds to the distribution of outcomes that would have prevailed in group 2 if individuals' characteristics X had remained as in group 1. Let $F_{H_2Y_2X_1U_2}$ denote the joint cumulative distribution of income and health in this case. The endowment effect $\Delta^X(C_A)$ can then be written as:

$$\Delta^X(C_A) = C_A(q; F_{H_2Y_2X_2U_2}) - C_A(q; F_{H_2Y_2X_1U_2}). \quad (1.11)$$

Let $F_{H_2Y_2X_1U_1}$ denote the joint cumulative distribution of income and health that would have occurred for group 2 had individual characteristics in that group been rewarded according to the return regime of group 1. The structural effect $\Delta^S(C_A)$ can then be written as:

$$\Delta^S(C_A) = C_A(q; F_{H_2Y_2X_1U_1}) - C_A(q; F_{H_1Y_1X_1U_1}). \quad (1.12)$$

¹See Fortin, Lemieux, and Firpo (2011)

These two last equations lead to:

$$\begin{aligned}
C_A(q; F_{H_2Y_2X_2U_2}) - C_A(q; F_{H_1Y_1X_1U_1}) &= \underbrace{[C_A(q; F_{H_2Y_2X_2U_2}) - C_A(q; F_{H_2Y_2X_1U_2})]}_{\text{endowment effect}} \\
&+ \underbrace{[C_A(q; F_{H_2Y_2X_1U_1}) - C_A(q; F_{H_1Y_1X_1U_1})]}_{\text{structural effect}} \\
&+ \underbrace{[C_A(q; F_{H_2Y_2X_1U_2}) - C_A(q; F_{H_2Y_2X_1U_1})]}_{\text{change in the distribution of unobservables}}. \quad (1.13)
\end{aligned}$$

To identify the above components, we rely on the conditional independence assumption, which states that the distribution of U is independent of G given $X = x$, i.e., $U \perp\!\!\!\perp G \mid X = x$. In the program evaluation literature, this assumption is sometimes called unconfoundedness or selection on observables. It may allow the researcher to provide a causal interpretation of the decomposition methodology's results and to identify the treatment effect parameter. However, if violated, the differences attributed to the endowment and structural effects could partly reflect changes in unobserved factors. Hence, the decomposition remains purely descriptive.

This assumption implies that $F_{U_1|X}(u \mid x) = F_{U_2|X}(u \mid x)$ in Equation (1.8). As a result, we obtain

$$C_A(q; F_{H_2Y_2X_1U_2}) = C_A(q; F_{H_2Y_2X_1U_1}),$$

which implies that $\Delta^U(C_A) = 0$. Hence, Equation (1.13) becomes:

$$\begin{aligned}
C_A(q; F_{H_2Y_2X_2U_2}) - C_A(q; F_{H_1Y_1X_1U_1}) &= \underbrace{[C_A(q; F_{H_2Y_2X_2U_2}) - C_A(q; F_{H_2Y_2X_1U_2})]}_{\text{endowment effect}} \\
&+ \underbrace{[C_A(q; F_{H_2Y_2X_1U_1}) - C_A(q; F_{H_1Y_1X_1U_1})]}_{\text{structural effect}}. \quad (1.14)
\end{aligned}$$

Finally, Equation (1.9) can be simplified to :

$$\begin{aligned}
C_A(q; F_{H_2Y_2X_2}) - C_A(q; F_{H_1Y_1X_1}) &= \underbrace{[C_A(q; F_{H_2Y_2X_2}) - C_A(q; F_{H_2Y_2X_1})]}_{\text{endowment effect}} \\
&+ \underbrace{[C_A(q; F_{H_2Y_2X_1}) - C_A(q; F_{H_1Y_1X_1})]}_{\text{structural effect}}. \quad (1.15)
\end{aligned}$$

One can also take group 1 as the anchor point. Then

$$\begin{aligned}
C_A(q; F_{H_2Y_2X_2}) - C_A(q; F_{H_1Y_1X_1}) &= \underbrace{[C_A(q; F_{H_2Y_2X_2}) - C_A(q; F_{H_1Y_1X_2})]}_{\text{endowment effect}} \\
&\quad + \underbrace{[C_A(q; F_{H_1Y_1X_2}) - C_A(q; F_{H_1Y_1X_1})]}_{\text{structural effect}}. \quad (1.16)
\end{aligned}$$

To be neutral on the choice of the anchor point, we use the average of the effects from the two anchor points. In this case,

$$\begin{aligned}
\Delta^O(C_A) &= \underbrace{0.5 * [C_A(q; F_{H_2Y_2X_2}) - C_A(q; F_{H_2Y_2X_1})] + 0.5 * [C_A(q; F_{H_2Y_2X_2}) - C_A(q; F_{H_1Y_1X_2})]}_{\text{endowment effect}} \\
&\quad + \underbrace{0.5 * [C_A(q; F_{H_2Y_2X_1}) - C_A(q; F_{H_1Y_1X_1})] + 0.5 * [C_A(q; F_{H_1Y_1X_2}) - C_A(q; F_{H_1Y_1X_1})]}_{\text{structural effect}}. \quad (1.17)
\end{aligned}$$

By analogy, we can retrieve the change in the concentration curves $\Delta^O(C_R)$ as follows:

$$\begin{aligned}
\Delta^O(C_R) &= C_R(q; F_{H_2Y_2X_2}) - C_R(q; F_{H_1Y_1X_1}) \\
&= \underbrace{0.5 * [C_R(q; F_{H_2Y_2X_2}) - C_R(q; F_{H_2Y_2X_1})] + 0.5 * [C_R(q; F_{H_2Y_2X_2}) - C_R(q; F_{H_1Y_1X_2})]}_{\text{endowment effect}} \\
&\quad + \underbrace{0.5 * [C_R(q; F_{H_2Y_2X_1}) - C_R(q; F_{H_1Y_1X_1})] + 0.5 * [C_R(q; F_{H_1Y_1X_2}) - C_R(q; F_{H_1Y_1X_1})]}_{\text{structural effect}}. \quad (1.18)
\end{aligned}$$

1.2.3 Reweighting approach

The reweighting is our first approach to decompose the change in both the absolute and relative concentration curves. This approach is typically used when one is interested in the overall decomposition effect, without specifying the individual effect of each covariate.

As previously indicated, the decomposition of the change requires constructing the counterfactual curves. These designate the curves that would have prevailed if individual characteristics had remained as in group 1, while individuals had been rewarded according to the return system of group 2.

We first define the counterfactual absolute concentration curve as follows:

$$C_A(q; F_{H_2Y_2X_1}) = \int_{\mathcal{X}} \int_0^{F_Y^{-1}(q)} \int_0^{h_{max}} h f_{H_2|YX}(h|y, x) dh f_{Y_2|X}(y|x) dy dF_{X_1}(x). \quad (1.19)$$

DiNardo, Fortin, and Lemieux (1996) suggest a reweighting function to construct the counterfactual densities, which can be easily adapted to construct the counterfactual absolute concentration curve. In fact, Equation (1.19) becomes:

$$C_A(q; F_{H_2Y_2X_1}) = \int_{\mathcal{X}} \int_0^{F_Y^{-1}(q)} \int_0^{h_{max}} h f_{H_2|YX}(h|y, x) dh f_{Y_2|X}(y|x) dy \Psi_x(X) dF_{X_2}(x), \quad (1.20)$$

where $\Psi_x(X) = dF_{X_1}(x)/dF_{X_2}(x)$ is the reweighting ratio. Using Bayes' rule, DiNardo, Fortin, and Lemieux (1996) show that this ratio becomes:

$$\Psi_x(X) = \frac{Pr(g = 1|X)}{Pr(g = 2|X)} \cdot \frac{Pr(g = 2)}{Pr(g = 1)}. \quad (1.21)$$

Hence, we can deduce Equation (1.15) by replacing the formula for $C_A(q; F_{H_2Y_2X_1})$.

Similarly, the counterfactual relative concentration curve is :

$$C_R(q; F_{H_2Y_2X_1}) = \frac{1}{\mu_{h2}} \int_{\mathcal{X}} \int_0^{F_Y^{-1}(q)} \int_0^{h_{max}} h f_{H_2|YX}(h|y, x) dh f_{Y_2|X}(y|x) dy \Psi_x(X) dF_{X_2}(x). \quad (1.22)$$

To estimate $\Psi_x(X)$, we can use a binary model (logit or probit) to predict the probability of belonging to groups 1 and 2. The method consists of pooling the data from groups 1 and 2 and running a binary model (logit or probit) on the probability of belonging to group 1, by creating a binary variable $D = 1$ if $g = 1$ and $D = 0$ if $g = 2$. Then, the reweighting factor $\Psi(x_i)$ for observations in group 2 can be estimated using the predicted probabilities $\widehat{Pr}[D = 1]$, $\widehat{Pr}[D = 0]$, $\widehat{Pr}[D = 1|X = x]$, and $\widehat{Pr}[D = 0|X = x] = 1 - \widehat{Pr}[D = 1|X = x]$. Then,

$$\widehat{\Psi}(x_i) = \frac{\widehat{Pr}[D = 1|X = x_i]/\widehat{Pr}[D = 1]}{\widehat{Pr}[D = 0|X = x_i]/\widehat{Pr}[D = 0]}. \quad (1.23)$$

$F_{H_2Y_2X_1}(h)$ is estimated using the observations of population of group 2 and the following estimator:

$$\widehat{F}_{H_2Y_2X_1}(h) = \frac{1}{N_2} \sum_{i=1}^{N_2} \mathbb{1}(y_i \leq h) \widehat{\Psi}(x_i), \quad (1.24)$$

where N_2 is the number of observations in the survey of the population of group 2. The counterfactual quantile function is obtained by inverting (1.24):

$$\widehat{F}_{H_2Y_2X_1}^{-1}(q) = \inf y : \widehat{F}_{H_2Y_2X_1}(y) \geq q. \quad (1.25)$$

1.2.4 Recentered influence function

To link the change in the distribution of individuals' characteristics to the change in the absolute and relative concentration curves, we build our methodology on the concept of the influence function (IF) introduced by Hampel (1968, 1974). The IF is defined as the influence of an individual observation on a distributional statistic of a marginal distribution of an outcome variable. Since the absolute and relative concentration curves are functionals of F_{HY} , we can define their influence functions.

This represents our second approach, which we use to decompose the change in the absolute and relative concentration curves. In contrast to the first approach, this method allows the decomposition of the individual effect of each covariate of interest. In what follows, we focus on the absolute concentration curve, but all results for the relative concentration curve can be derived similarly.

Thus, the influence function of the absolute concentration curve $C_A(q; F_{HY})$ is its first-order directional derivative.

Let δ_{h_i, y_i} denote the Dirac distribution of (H, Y) with a probability mass of 1 at (h_i, y_i) , and let $\widetilde{f}_{HY}^t = t \cdot \delta_{h_i, y_i} + (1-t) \cdot f_{HY}$. The influence function of the coordinate of the absolute concentration curve at q is given by

$$IF(h_i, y_i; C_A(q; F_{HY})) = \left. \frac{\partial}{\partial t} C_A(q; \widetilde{f}_{HY}^t) \right|_{t=0}. \quad (1.26)$$

In their paper, Firpo, Fortin, and Lemieux (2009) define the *recentered influence function (RIF)* as the distributional statistic added to its influence function. They show that the RIF corresponds to the leading term of the Mises (1947) linear approximation of the distributional statistic. Following the same logic, we define the RIF of the absolute concentration curve $C_A(q; F_{HY})$ as

$$RIF(h_i, y_i; C_A(q; F_{HY})) = C_A(q; F_{HY}) + IF(h_i, y_i; C_A(q; F_{HY})). \quad (1.27)$$

Propositions 1 and 2 provide the expressions of the influence functions for the absolute and relative concentration curves, respectively. The proofs of both propositions are provided in the appendix.

Proposition 1. *The influence function of the coordinate $C_A(q; F_{HY})$ of the absolute health concentration curve is given by*

$$\begin{aligned} IF(h_i, y_i; C_A(q; F_{HY})) &= -C_A(q; F_{HY}) + [q - \mathbb{1}(y_i \leq F_Y^{-1}(q))] \cdot E[H | Y = F_Y^{-1}(q)] \\ &\quad + \mathbb{1}(y_i \leq F_Y^{-1}(q)) \cdot h_i. \end{aligned}$$

Its recentered influence function is given by

$$RIF(h_i, y_i; C_A(q; F_{HY})) = [q - \mathbb{1}(y_i \leq F_Y^{-1}(q))] \cdot E[H | Y = F_Y^{-1}(q)] + \mathbb{1}(y_i \leq F_Y^{-1}(q)) \cdot h_i$$

Proposition 2. *The influence function of the coordinate $C_R(q; F_{HY})$ of the health concentration curve is given by*

$$\begin{aligned} IF(h_i, y_i; C_R(q; F_{HY})) &= -\frac{C_R(q; F_{HY})}{\mu_H} \cdot h_i + [q - \mathbb{1}(y_i \leq F_Y^{-1}(q))] \cdot \frac{E[H | Y = F_Y^{-1}(q)]}{\mu_H} \\ &\quad + \mathbb{1}(y_i \leq F_Y^{-1}(q)) \cdot \frac{h_i}{\mu_H}. \end{aligned}$$

Its recentered influence function is given by

$$\begin{aligned} RIF(h_i, y_i; C_R(q; F_{HY})) &= C_R(q; F_{HY}) \cdot \left[1 - \frac{h_i}{\mu_H} \right] + [q - \mathbb{1}(y_i \leq F_Y^{-1}(q))] \cdot \frac{E[H|Y = F_Y^{-1}(q)]}{\mu_H} \\ &\quad + \mathbb{1}(y_i \leq F_Y^{-1}(q)) \cdot \frac{h_i}{\mu_H}. \end{aligned}$$

Since

$$E(IF(h_i, y_i; C_A(q; F_{HY}))) = \int_0^\infty \int_0^{h_{\max}} IF(h_i, y_i; C_A(q; F_{HY})) dF_{H|Y}(h|y) dF_Y(y) = 0,$$

Equation (1.27) implies that

$$E(RIF(h_i, y_i; C_A(q; F_{HY}))) = C_A(q; F_{HY}).$$

Thus, in the presence of covariates X , we can use the law of iterated expectations to express the unconditional expectation as

$$\begin{aligned} E(RIF(h_i, y_i; C_A(q; F_{HY}))) &= E\left(E(RIF(h_i, y_i; C_A(q; F_{HY})) | X)\right) \\ &= \int E(RIF(h_i, y_i; C_A(q; F_{HY})) | X) dF_X. \end{aligned}$$

The conditional expectation of the RIF can then be estimated using regression methods. In this paper, we adopt a linear specification in which the conditional expectation is a linear function of the observed covariates:

$$E(RIF(h_i, y_i; C_A(q; F_{HY})) | X) = X\beta, \quad E(RIF(h_i, y_i; C_R(q; F_{HY})) | X) = X\gamma,$$

assuming that the expected value of the approximation error is zero (see Firpo, Fortin, and Lemieux (2009)).

Using OLS, we estimate :

$$RIF(h_i, y_i; C_A(q, F_{Y_g, H_g})) = X_g \beta_g + \epsilon_g \quad g = 1, 2, \quad (1.28)$$

and

$$RIF(h_i, y_i; C_R(q, f_{Y_g, H_g})) = X_g \gamma_g + \epsilon_g \quad g = 1, 2. \quad (1.29)$$

Assuming a linear specification for the RIF regression model allows us to further decompose the endowment and structural effects into the contributions of the relevant covariates.

1.3 Empirical illustration

1.3.1 Data and descriptive statistics

To illustrate the approaches described above, we use U.S. data on body mass index (BMI) and cigarette consumption from the National Health Interview Survey (NHIS). First conducted in July 1957, the NHIS aims to monitor the health of the civilian noninstitutionalized population of the United States. It is a repeated cross-sectional survey and provides an important source of information on trends in illness, disability, and the utilization of related health services.

To compare changes in socioeconomic health inequalities between 2000 and 2020, we focus on two indicators: cigarette consumption and overweightness. The first measures the number of cigarettes smoked per day, ranging from 0 for non-smokers to a maximum of 95 cigarettes. The second relies on BMI. We use the measure $\text{Max}[0, BMI - 25]$ developed by Bilger, Kruger, and Finkelstein (2017) to reflect how far an individual's BMI exceeds the threshold for "normal" weight.

To determine socioeconomic ranks, we use family income adjusted for family size by dividing total family income by the number of household members. Respondents were asked whether any family member received income from various sources (e.g., wages, salary from self-employment, Social Security, railroad retirement, government assistance) and to provide their best estimate of total family income before taxes for the previous calendar year. Individual ranks were calculated separately for each survey year.

We use publicly available data restricted to the adult population, yielding 31,410 observations for cigarette consumption and 28,982 for overweightness in 2000, and

30,007 observations for cigarette consumption and 27,834 for overweightness in 2020.

Table 1.1 presents descriptive statistics for the variables used in this illustration. It shows that overweightness has increased over the 20-year period: the mean of the $\text{Max}[0, \text{BMI} - 25]$ indicator rose from 2.6 Kg/m^2 in 2000 to 3.7 Kg/m^2 in 2020. In addition, the average BMI for the full sample increased from 26.4 Kg/m^2 to 28.7 Kg/m^2 , and the proportion of individuals with obesity ($\text{BMI} \geq 30$) increased from 20.6% in 2000 to 30.3% in 2020.

Conversely, cigarette consumption declined over the same period. Table 1.1 shows that the average number of cigarettes smoked per day (including non-smokers) fell from 3.4 in 2000 to 1.3 in 2020. Moreover, the proportion of daily smokers decreased by nearly half, from 41.9% in 2000 to 23.4% in 2020.

In terms of inequality, Figure 1.1 plots the relative concentration curves for overweightness and cigarette consumption for 2000 and 2020. In both panels, all curves lie above the 45-degree line, indicating pro-poor inequalities: ill health, as measured by overweightness and cigarette consumption, is concentrated among poorer individuals. While the right panel clearly shows the dominance of the 2000 relative concentration curve for cigarette consumption over the 2020 curve, the dominance of the curves for overweightness is less visually evident. The results of the dominance tests at the top of Table 1.2 confirm that 2000 dominates 2020 for both indicators. This suggests that, from a social perspective, health inequalities were more favorable (pro-poor) in 2000 than in 2020.

Regarding health shortfalls, Figure 1.2 presents the absolute concentration curves for overweightness and cigarette consumption in 2000 and 2020. The figure reveals a clear dominance of the 2000 curve over the 2020 curve for overweightness, and a dominance of the 2020 curve over the 2000 curve for cigarette consumption. These patterns are corroborated by the dominance tests reported at the bottom of Table 1.2. As previously noted, the right panel of Figure 1.2 illustrates a decline in cigarette consumption across all socioeconomic groups between 2000 and 2020. In contrast, overweightness increased over the same period, as shown in the left panel.

1.3.2 Distributional decomposition of changes in cigarette consumption

In this subsection, we present the results of the distributional decomposition of changes in cigarette consumption. We first focus on decomposing the relative and absolute concentration curves using the reweighting approach. We then present the results obtained from decomposing these two curves using RIF regressions.

Cigarette use depends on a range of demographic and environmental factors, some of which are modifiable by individuals or their environment—such as education, region of residence, and marital status—and others that are not, including age, sex, and race (Bondurant and Wedge, 2009; CDC, 2024). Two models are considered in this analysis. The outcome variable is the average number of cigarettes smoked per day, including zero for non-smokers. In our setting, the distributional statistic is not the mean of the outcome variable, but rather the relative and the absolute concentration curves of cigarette consumption. In fact, various factors affect the joint distribution of health and income and, in turn, its distributional statistics.

In the first model, the covariates include non-choice variables: sex, age, and four race dummy variables. In the second model, we additionally include choice variables, namely education (five dummy variables) and region of residence (four dummy variables). Finally, we add a dummy variable indicating whether the household has children, as parenthood may influence both smoking prevalence and smoking behavior (Jarvis, 1996; Johansson, Halling, and The LinQuest Study Group, 2003). We therefore exclude marital status due to its high collinearity with the presence of children in the household.

Our analysis accounts for the complex NHIS sampling design by incorporating survey weights constructed through multiple adjustment stages, beginning with the inverse probability of selection and adjusted for nonresponse and post-stratification. This ensures national representativeness. We adopt linear functional forms for key covariates, which are commonly used in the literature and facilitate interpretation. Finally, height, weight, and cigarette use are self-reported in the NHIS and may be subject to measurement error due to misreporting or social desirability bias. Such errors are

likely to attenuate estimated associations, suggesting that our results may represent conservative estimates.

	Outcome variable	Covariates
Model 1	Cigarettes per day	Sex, Age, Race
Model 2	Cigarettes per day	Sex, Age, Race, Region, Education, Children (Y/N)

Reweighted relative concentration curves

Counterfactual curves

Following the approach developed in Section 1.2.3, we estimate counterfactual concentration curves for 2000 and 2020 under both models. Figure 1.3 presents the relative concentration curves and their counterfactuals for Model 2, while Figure 1.4 displays the counterfactual relative concentration curves for both models.

In Figure 1.3, the right panel shows a clear dominance of the observed relative concentration curve for 2000 over its counterfactual constructed using the returns of 2020. Interestingly, this counterfactual curve even lies above the observed relative concentration curve for 2020. Conversely, the counterfactual curve for 2020, constructed using the returns of 2000, lies below both the observed relative concentration curves for 2000 and 2020. These visual patterns are confirmed by the dominance tests reported in Table 1.3.

When comparing the two models, Figure 1.4 reveals no substantial differences between the counterfactual relative concentration curves. Specifically, in the right panel, the counterfactual curves for 2000 constructed using the returns of 2020 under Models 1 and 2 almost completely overlap across the entire socioeconomic rank distribution. In the left panel, the counterfactual curve for 2020 constructed using the returns of 2000 under Model 2 lies slightly above the corresponding curve under Model 1 around the 62nd percentile; however, the two curves converge again across the remaining ranks. No formal dominance tests were conducted to assess the statistical significance of these differences.

Overall, the visual evidence suggests that the counterfactual curve for a given year closely resembles the observed curve of the other year once the returns of that year are imposed. This finding implies that a substantial share of the observed change

in relative concentration curves between 2000 and 2020 is likely attributable to the unexplained component, which we examine in greater detail in the following section.

Endowment and structural effects

The endowment and structural effects are constructed using the counterfactual relative concentration curves described above. The corresponding curves can be written as:

$$\text{Endowment}_{C_R} = 0.5 * (C_R(2020) - \text{Counter}C_R(2000)) + 0.5 * (\text{Counter}C_R(2020) - C_R(2000)), \quad (1.30)$$

and

$$\text{Structural}_{C_R} = 0.5 * (\text{Counter}C_R(2000) - C_R(2000)) + 0.5 * (C_R(2020) - \text{Counter}C_R(2020)). \quad (1.31)$$

As shown in Figure 1.5, the structural effect is positive and larger in absolute value than the endowment effect. In contrast, the endowment effect is negative and close to zero—often indistinguishable from zero—across much of the socioeconomic rank distribution. Moreover, both models yield very similar results, particularly for the structural effect, for which the curves fully overlap. In contrast, the endowment effect curves intersect at several points along the socioeconomic rank, with no clear dominance. Overall, these findings suggest that the changes observed in the relative concentration curves between 2000 and 2020 are primarily driven by the structural effect, which shifts the curves upward and increases inequality.

Reweighted absolute concentration curves

Counterfactual curves

Similarly to the relative concentration curves, we estimate counterfactual absolute concentration curves for 2000 and 2020 under both models. Figure 1.6 presents the absolute concentration curves and their counterfactuals for Model 2, while Figure 1.7 displays the counterfactual absolute concentration curves for both models.

Figure 1.6 reveals a clear dominance pattern. In the right panel, the counterfactual absolute concentration curve for 2000, constructed using the returns of 2020, lies

entirely below the observed curve for 2000 and entirely above the observed curve for 2020. Conversely, the counterfactual curve for 2020 constructed using the returns of 2000 lies above the observed curve for 2020 and below the observed curve for 2000. These graphical patterns are confirmed by the dominance tests reported in Table 1.4. Figure 1.7 compares the counterfactual curves across the two models, where differences become more apparent. In the right panel, the counterfactual curve for 2000 constructed using the returns of 2020 under Model 1 lies very close to the observed curve for 2020. However, when additional covariates are included, as in Model 2, the absolute concentration curve shifts upward, suggesting that the change between 2000 and 2020 may be partially attributable to the endowment effect. A similar pattern emerges in the left panel of Figure 1.7, where the counterfactual curve for 2020 constructed using the returns of 2000 under Model 2 shifts downward away from the observed 2000 curve once additional covariates are included.

Endowment and structural effects

The endowment and the structural effects are constructed using the previously defined counterfactual absolute concentration curves. The curves of the endowment and structural effects can be written as:

$$Endowment_{C_A} = 0.5 * (C_A(2020) - CounterC_A(2000)) + 0.5 * (CounterC_A(2020) - C_A(2000)), \quad (1.32)$$

and

$$Structural_{C_A} = 0.5 * (CounterC_A(2000) - C_A(2000)) + 0.5 * (C_A(2020) - CounterC_A(2020)). \quad (1.33)$$

As shown in Figure 1.8, both the endowment and structural effects are negative, reflecting the overall reduction in cigarette consumption between 2000 and 2020. However, the structural effect is larger in absolute value than the endowment effect. Notably, when additional covariates are included in the model, the structural effect decreases in absolute value in favour of the endowment effect, indicating that observed characteristics capture a larger share of the change in cigarette consumption

over this period. In the following section, we examine the factors driving this change in greater detail.

RIF regression for the relative concentration curves

As described in Subsection 1.2.4, the RIF of the concentration curve for cigarette consumption is constructed and plotted in the left panel of Figure 1.9. This panel compares the observed concentration curve with the expected RIF-based concentration curve. The figure shows that the two curves are very close, indicating a good approximation.

In addition, to assess the fit of the RIF model, we apply an out-of-sample validation procedure. Specifically, the sample is randomly split into two subsamples: a training subsample comprising 80% of the observations, used to estimate the model, and a validation subsample consisting of the remaining 20%. The following regression is estimated on the training subsample:

$$RIF_{CR}^q = X_i\beta + \epsilon_i.$$

Using the validation subsample, we then construct the counterfactual relative concentration curve, $E[\widehat{RIF}]$, and compare it with the observed concentration curve computed on the same validation sample. The right panel of Figure 1.9 illustrates this comparison: the two curves are very similar, and their confidence intervals overlap substantially, providing further evidence of the good performance of the RIF specification.

Endowment and structural effects

As expected, adding covariates to Model 2 increases the difference between the endowment effect and the structural effect. In both models, the structural effect is positive and greater in absolute value than the endowment effect, which is negative. In both cases, the change observed in the concentration curves between 2000 and 2020 is partially driven by the structural effect.

Endowment and structural effects of each variable

The use of the Oaxaca–Blinder decomposition, combined with the RIF framework, allows us to assess the contribution of each variable or factor to cigarette consumption. Figure 1.11 plots and compares the contributions of each variable in Model 2 to the endowment and structural effects.

An examination of these figures shows that the main contributors to the endowment effect are education and the presence of children in the household. Both factors, through their negative endowment effects across the entire socioeconomic distribution, contribute to a reduction in inequality in cigarette consumption. In addition, gender and region of residence, although their contributions are relatively small, also act to reduce inequality. By contrast, age and race exhibit heterogeneous patterns across the socioeconomic rank distribution: the endowment effect of age changes sign around the middle of the distribution, whereas the effect of race is negative and significant only at the top of the distribution.

Turning to the structural effect, the overall impact is largely driven by the intercept, which is positive and sizeable, thereby determining the sign of the total structural effect. Except for education, the structural effects of most covariates are either statistically insignificant or small and positive only at the top of the distribution. The structural effect of education remains negative and relatively large in absolute value; however, it is not sufficient to offset the dominant contribution of the intercept.

Overall, the endowment effect is mainly driven by education and the presence of children in the household, both of which contribute to a reduction in inequality in cigarette consumption. In contrast, the structural effect is largely dominated by the intercept, which explains most of the increase in inequality, while the contributions of other covariates are generally small or insignificant.

RIF regression for the absolute concentration curves

A similar approach is employed for the absolute concentration curve to assess the fit of the RIF model. The expected RIF-based curve is compared with the observed absolute concentration curve in the left panel of Figure 1.12. In addition, the counterfactual

curve constructed using the validation subsample is compared with its observed counterpart in the right panel of Figure 1.12. In both cases, the expected and observed curves closely align, indicating a good model fit.

Endowment and structural effects

Figure 1.13 presents the results for both models, showing that both the endowment and structural effects are negative. However, when additional covariates are included, the magnitude of the structural effect decreases and shifts toward the endowment effect, suggesting that the added variables capture a larger share of the observed change in cigarette consumption shortfall.

Endowment and structural effects of each variable

The contributions of individual demographic factors for model 2 are displayed in Figure 1.14. An examination of the endowment effects indicates that education remains the main contributor to the reduction in cigarette consumption shortfall. In addition, age and the presence of children in the household reduce the shortfall, whereas gender and race tend to increase it.

Turning to the structural effect, changes in the returns to demographic characteristics are largely driven by the intercept. The structural effect of the intercept is negative and large in absolute value, suggesting that changes in the broader social structure—beyond the demographic characteristics explicitly included in the model—play a major role in reducing the cigarette consumption shortfall. While the returns to age, gender, race, and education have a positive effect on the shortfall, they do not offset the dominant influence of the intercept and the presence of children in the household.

Overall, the RIF-based decomposition of the absolute and relative concentration curves confirms that the decline in cigarette consumption between 2000 and 2020 is driven by both endowment and structural effects. At the variable level, improvements in education and changes in household composition, particularly the presence of children, play a key role in reducing cigarette consumption shortfalls, whereas the returns to most demographic characteristics have a more limited association. These results highlight the importance of broader structural changes in explaining long-term trends in smoking behaviour.

1.3.3 Distributional decomposition of change in overweightness

Following the same approaches applied to cigarette consumption, we use the overweightness indicator described in subsection 1.3.1 as the outcome variable, along with the same covariates as in subsection 1.3.2. However, only the results of Model 2 are briefly discussed in this section.

Reweighted absolute and relative concentration curves

The counterfactual absolute and relative concentration curves, obtained using the reweighting approach, are shown in Figure 1.15.

Regarding the concentration curves at the top of the figure, the differences between the relative concentration curves of 2020 and 2000, and their corresponding counterfactuals, are not visually obvious. Nonetheless, the results of the dominance tests presented in Table 1.5 reveal certain dominance relationships between the curves, indicating that the change in socioeconomic inequality in overweightness remains limited. This finding is further confirmed by the small magnitude or insignificance of the endowment and structural effects plotted on the left side of Figure 1.16. While the structural effect appears strictly positive in the middle of the social rank distribution, the endowment effect is mostly insignificant across the social rank.

Considering the change in overweightness shortfall, as expressed by the absolute concentration curves at the bottom of Figure 1.15, the counterfactual curves of 2020 and 2000 are very close to the actual curves of 2000 and 2020, respectively. This suggests that the observed change in the absolute concentration curves between 2020 and 2000 is largely driven by the structural effect. However, the dominance tests in Table 1.6—which show dominance of the actual 2020 curve over the counterfactual 2000 curve, and dominance of the counterfactual 2020 curve over the actual 2000 curve—indicate a small but significant endowment effect. This is visually reflected in the right panel of Figure 1.16.

The change in socioeconomic inequality in overweightness between 2000 and 2020 is largely driven by structural effects, with the endowment effect being small but occasionally significant. Relative concentration curves show little visual change, while

absolute concentration curves confirm that structural factors primarily explain the observed differences. Overall, the evolution of overweightness inequality remains limited over time.

RIF regression for the absolute and the relative concentration curves

The results of the decomposition of changes in the absolute and relative concentration curves using RIF regression are presented in Figures 1.18 and 1.19.

The left panel of Figure 1.18 plots the structural and the endowment effects for the relative concentration curve. It shows a positive and significant structural (return) effect for Model 2, indicating that changes in the returns to characteristics increase socioeconomic inequalities in overweightness. The contribution of individual covariates is mostly insignificant (bottom panel of Figure 1.19). In contrast, the negative and significant endowment effect shifts the concentration curve downward, thereby reducing inequalities. Education and the presence of children in the household are the main contributors to the endowment effect, helping to reduce inequality. At the same time, age contributes positively at the top of the distribution, increasing socioeconomic disparities (top panel of Figure 1.19).

The right panel of Figure 1.18 shows the decomposition of the absolute concentration curve into endowment and structural effects. In contrast, Figure 1.20 displays the contribution of each demographic characteristic to these effects. These figures indicate that the change in the socioeconomic shortfall of overweightness is mainly driven by the structural effect, which is larger in absolute value than the endowment effect. The structural effect of the intercept is the dominant contributor to the overall change in returns to demographic characteristics (bottom panel of Figure 1.20). Regarding the endowment effect, education remains the largest factor reducing the overweightness shortfall, followed by the presence of children, whereas age contributes positively except at the bottom of the distribution, reflecting the higher prevalence of overweightness among an ageing population.

The RIF regression approach confirms the results from the reweighting approach, showing that changes in socioeconomic inequality in overweightness are primarily

driven by the structural effect, which increases disparities. In contrast, the endowment effect is smaller and generally reduces inequalities. Education and the presence of children in the household are the main factors lowering inequality, whereas age contributes positively at the top of the distribution, reflecting higher overweightness among older individuals. Overall, structural factors dominate the evolution of overweightness inequality and shortfall over time.

1.4 Conclusion

A large body of literature on health inequality has emerged in recent decades. Wagstaff, Paci, and Doorslaer (1991) examined several tools used to measure health inequalities. One of the most widely used tools is the concentration index and its associated concentration curves, which measure socioeconomic inequalities in health.

In this paper, we propose a framework to decompose both absolute and relative concentration curves. Our approach extends previous methods in the literature, which have typically focused on index decomposition, by enabling the decomposition of the entire curve. To illustrate this framework, we use data on body mass index (BMI) and cigarette consumption from the National Health Interview Survey (NHIS). The reweighting approach provides an aggregate decomposition of changes in inequality, while the RIF regression approach allows a detailed decomposition, quantifying each factor's contribution.

Our results show that changes in socioeconomic inequality in overweightness and cigarette consumption are primarily driven by structural effects, which tend to increase disparities. In contrast, endowment effects are smaller and often act to reduce inequalities. Education and the presence of children in the household are the main factors mitigating inequality, whereas age contributes positively, reflecting higher risks among older populations. Overall, structural factors dominate the evolution of health inequalities over time, highlighting the importance of broader social and environmental determinants in shaping health outcomes.

Although the results are mostly uniform across the socioeconomic rank distribution—except for the contribution of age, which changes sign—our approach relies on dominance comparisons of relative and absolute concentration curves and therefore provides a more robust ranking of health inequalities that is independent of index specification. In contrast, index-based measures, such as the concentration index, the health achievement index, or widely used summary statistics like the mean, may yield comparisons that depend on the chosen index and can lead to contradictory conclusions across distributions. Furthermore, Chapter 2 provides a clear example in which the average treatment effect is not statistically significant, while the effect is

significant below the 85th socioeconomic rank.

Although these decomposition methods help in understanding the main determinants of changes in inequality, they do not provide a causal interpretation of the results. With the growing literature on program evaluation and treatment effects, future research could explore the relationship between structural effects and causal treatment effects, providing further insight into the mechanisms underlying health inequalities.

Appendix

Proof of Proposition 1: Let δ_{h_i, y_i} be the bivariate distribution of (H, Y) with a probability mass of 1 at (h_i, y_i) , and let $\tilde{f}_{HY}^t = t \cdot \delta_{h_i, y_i} + (1-t) \cdot f_{HY}$ and . The influence function of the coordinate of the generalized health concentration curve at q is given by

$$IF(h_i, y_i; C_A(q; F_{HY})) = \left. \frac{\partial}{\partial t} C_A(q; \tilde{f}_{HY}^t) \right|_{t=0} \quad (1.34)$$

It is useful to rewrite equation (1.2) as

$$C_A(q; F_{HY}) = \int_0^{F_Y^{-1}(q)} \int_{\underline{h}}^{\bar{h}} h f_{H|Y}(h|Y=y) dh f_Y(y) dy \quad (1.35)$$

Using equation (1.35), equation (1.34) can be rewritten as

$$IF(h_i, y_i; C_A(q; F_{HY})) = \left. \frac{\partial}{\partial t} \left\{ \int_0^{\tilde{F}_Y^{-1}(q)} \int_0^{h_{\max}} h [t \cdot \delta_{h_i, y_i}(h, y) + (1-t) \cdot f_{HY}(hy)] dh dy \right\} \right|_{t=0}, \quad (1.36)$$

$$\begin{aligned} IF(h_i, y_i; C_A(q; F_{HY})) &= IF(y_i; F_Y^{-1}(q)) \cdot \int_0^{h_{\max}} h f_{HY}(h, F^{-1}(q)) dh \\ &\quad + \int_0^{F_Y^{-1}(q)} \int_0^{h_{\max}} h \delta_{h_i, y_i}(h, y) dh dy \\ &\quad - \int_0^{F_Y^{-1}(q)} \int_0^{h_{\max}} h f_{HY}(hy) dh dy. \end{aligned} \quad (1.37)$$

Note that $\int_0^{h_{\max}} h f_{HY}(h, F^{-1}(q)) dh = \int_0^{h_{\max}} h f_{H|Y}(h|Y=F^{-1}(q)) dh f_Y(F^{-1}(q))$. Also $\int_0^{h_{\max}} h f_{H|Y}(h|Y=F^{-1}(q)) dh = E[H|Y=F^{-1}(q)]$. From Firpo, Fortin, and Lemieux (2009), we know that $IF(y_i; F_Y^{-1}(q)) = [p - \mathbb{1}(y_i \leq F_Y^{-1}(q))] / f_Y(F_Y^{-1}(q))$. The first term on the r.h.s. of equation (1.37) becomes

$$IF(y_i; F_Y^{-1}(q)) \cdot \int_0^{h_{\max}} h f_{HY}(h, F^{-1}(q)) dh = [q - \mathbb{1}(y_i \leq F_Y^{-1}(q))] \cdot E[H|Y=F_Y^{-1}(q)]. \quad (1.38)$$

The second term on the r.h.s. of equation (1.37) integrates over a Dirac function with probability mass 1 at (h_i, y_i) . This implies that this term will be equal to h_i if $y_i \leq F_Y^{-1}(q)$ and equal to 0 if $y_i > F_Y^{-1}(q)$. The second term on the r.h.s. of equation (1.37) becomes

$$\int_0^{F_Y^{-1}(q)} \int_0^{h_{\max}} h \delta_{h_i, y_i}(h, y) dh dy = \mathbb{1}(y_i \leq F_Y^{-1}(q)) \cdot h_i. \quad (1.39)$$

Using equations (1.38) and (1.39) and the fact that the last term on the r.h.s. of equation (1.37) is exactly the expression of the $C_A(q; F_{HY})$ in equation (1.2), equation (1.37) becomes

$$\begin{aligned} IF(h_i, y_i; C_A(q; F_{HY})) &= [q - \mathbb{1}(y_i \leq F_Y^{-1}(q))] \cdot E[H|Y = F_Y^{-1}(q)] \\ &\quad + \mathbb{1}(y_i \leq F_Y^{-1}(q)) \cdot h_i - C_A(q; F_{HY}) \end{aligned} \quad (1.40)$$

Reorganizing the terms of the r.h.s. of equation (1.40) yields the first result of the proposition. Adding the functional $C_A(q; F_{HY})$ to the influence function yields the second result of the proposition. \square

Proof of Proposition 2: Since $C_R(q; F_{HY}) = C_A(q; F_{HY})/\mu_H$,

$$IF(h_i, y_i; C_R(q; F_{HY})) = \frac{IF(h_i, y_i; C_A(q; F_{HY})) \cdot \mu_H - C_A(q; F_{HY}) \cdot IF(h_i, y_i; \mu_H)}{\mu_H^2}. \quad (1.41)$$

The influence function of the average is a known result: $IF(h_i, y_i; \mu_H) = h_i - \mu_H$ (see Essama-Nssah and Lambert (2012)). Substituting this result and the expression of $IF(h_i, y_i; C_A(q; F_{HY}))$ into equation (1.41) yields

$$\begin{aligned} IF(h_i, y_i; C_R(q; F_{HY})) &= \frac{1}{\mu_H} [q - \mathbb{1}(y_i \leq F_Y^{-1}(q))] \cdot E[H|Y = F_Y^{-1}(q)] \\ &\quad + \mathbb{1}(y_i \leq F_Y^{-1}(q)) \cdot \frac{h_i}{\mu_H} - C_R(q; F_{HY}) \\ &\quad - C_R(q; F_{HY}) \cdot \frac{h_i}{\mu_H} + C_R(q; F_{HY}). \end{aligned} \quad (1.42)$$

Reorganizing the terms of the r.h.s. of equation (1.42) yields the first result of the proposition. Adding the functional $C_R(q; F_{HY})$ to the influence function yields the second result of the corollary. \square

TABLE 1.1: Descriptive statistics

	Year		
	2000	2020	Total
BMI			
Mean	26.44	27.83	27.12
Standard Deviation	4.82	5.50	5.21
Overweightness BMI-25			
Mean	2.62	3.72	3.16
Standard Deviation	3.71	4.57	4.19
Number of cigarettes smoked per day			
Mean	3.40	1.30	2.37
Standard Deviation	8.21	4.70	6.81
Per capita income			
Mean	20,808	40,940	30,644
Standard Deviation	18,444	33,673	28,796
Per capita income (2020 prices)			
Mean	27,372	40,940	
Standard Deviation	22,513	33,673	
Household region			
Northeast	18.56	17.65	18.12
Midwest	23.25	22.95	23.10
South	35.98	34.29	35.15
West	22.21	25.11	23.63
High level of education			
Drop out	23.09	9.46	16.41
High school degree or some college	45.96	37.03	41.59
Associate degree	8.68	13.21	10.90
Bachelor's degree	14.44	24.05	19.14
Graduate	7.83	16.25	11.95
Marital status			
Married or living with a partner	51.08	51.78	51.42
Widowed	10.70	11.01	10.85
Divorced or separated	16.58	16.79	16.68
Never married	21.64	20.41	21.04
Sex of Sample Adult			
Male	44.17	46.67	45.39
Female	55.83	53.33	54.61
Presence of children in the household			
No	62.40	74.29	68.23
Yes	37.60	25.71	31.77
Race groups with Hispanic origin			
Hispanic	16.35	11.90	14.17
Non-Hispanic White	66.67	71.04	68.81
Non-Hispanic Black	13.73	9.59	11.70
Non-Hispanic Other	3.25	7.48	5.32
Number of observations	31,410	30,007	61,417

Source of data: NHIS 2000 and NHIS 2020.

TABLE 1.2: Dominance tests

	p-Value	
	BMI	Cigarettes
$C_R(2000) \geq C_R(2020)$	1	1
$C_R(2020) \geq C_R(2000)$	0	0
$C_A(2000) \geq C_A(2020)$	1	0
$C_A(2020) \geq C_A(2000)$	0	0.9494949

Note: $C_R(2020)$ and $C_R(2000)$ refer to the relative concentration curves of 2020 and 2000, respectively, and $C_A(2020)$ and $C_A(2000)$ refer to the absolute concentration curves of 2020 and 2000, respectively.

TABLE 1.3: Dominance tests for counterfactuals of the relative concentration curves

	p-Value	
	Model 1	Model 2
$CounterC_R(2000) \geq C_R(2000)$	1	0.989899
$C_R(2000) \geq CounterC_R(2000)$	0	0
$CounterC_R(2000) \geq C_R(2020)$	1	0.8080808
$C_R(2020) \geq CounterC_R(2000)$	0	0.9494949
$CounterC_R(2020) \geq C_R(2020)$	0	0
$C_R(2020) \geq CounterC_R(2020)$	1	1
$CounterC_R(2020) \geq C_R(2000)$	0	0
$C_R(2000) \geq CounterC_R(2020)$	1	0.8080808

Note: $CounterC_R(2020)$ and $CounterC_R(2000)$ refer to the counterfactual of the relative concentration curve of 2020 with return of 2000 and the counterfactual of the relative concentration curve of 2000 with return of 2020, respectively.

TABLE 1.4: Dominance tests for counterfactuals of the absolute concentration curves

	p-Value	
	Model 1	Model 2
$CounterC_A(2000) \geq C_A(2000)$	0	0
$C_A(2000) \geq CounterC_A(2000)$	0.979798	0.9494949
$CounterC_A(2000) \geq C_A(2020)$	1	0.989899
$C_A(2020) \geq CounterC_A(2000)$	0	0
$CounterC_A(2020) \geq C_A(2020)$	0.969697	0.9494949
$C_A(2020) \geq CounterC_A(2020)$	0	0
$CounterC_A(2020) \geq C_A(2000)$	0	0
$C_A(2000) \geq CounterC_A(2020)$	1	0.989899

Note: $CounterC_A(2020)$ and $CounterC_A(2000)$ refer to the counterfactual of the generalized concentration curve of 2020 with return of 2000 and the counterfactual of the concentration curve of 2000 with return of 2020, respectively.

TABLE 1.5: Dominance tests for counterfactuals of the relative concentration curves of overweightness

	p-Value	
	Model 1	Model 2
$CounterC_R(2000) \geq C_R(2000)$	0.6060606	0.3636364
$C_R(2000) \geq CounterC_R(2000)$	0	0
$CounterC_R(2000) \geq C_R(2020)$	0.7373737	0.01010101
$C_R(2020) \geq CounterC_R(2000)$	0.8080808	0.6666667
$CounterC_R(2020) \geq C_R(2020)$	0	0
$C_R(2020) \geq CounterC_R(2020)$	0.959596	0.9292929
$CounterC_R(2020) \geq C_R(2000)$	0.5757576	0.2323232
$C_R(2000) \geq CounterC_R(2020)$	0.7373737	0.01010101

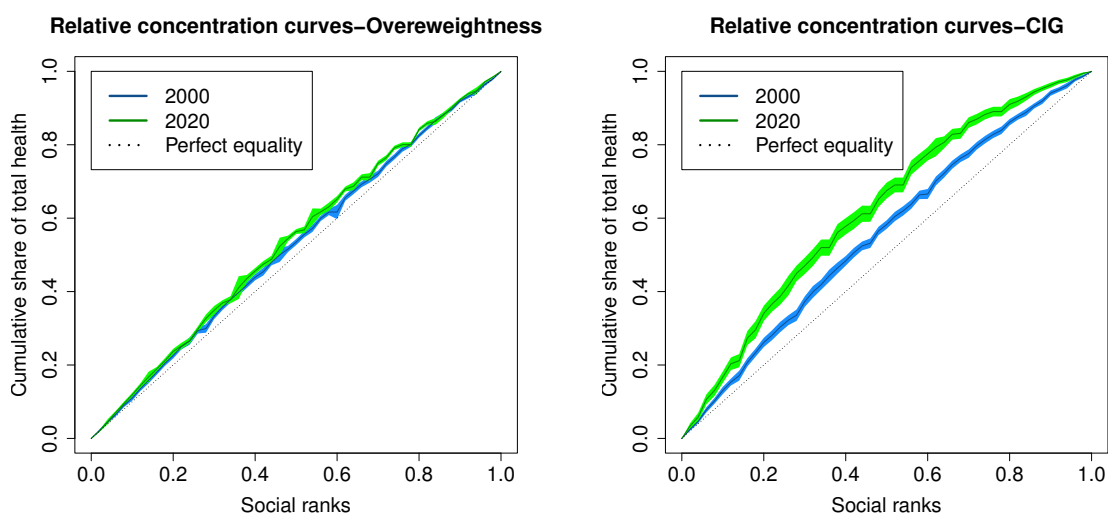
Note: $CounterC_R(2020)$ and $CounterC_R(2000)$ refer to the counterfactual of the relative concentration curve of 2020 with return of 2000 and the counterfactual of the relative concentration curve of 2000 with return of 2020, respectively.

TABLE 1.6: Dominance tests for counterfactuals of the absolute concentration curves of overweightness

	p-Value	
	Model 1	Model 2
$CounterC_A(2000) \geq C_A(2000)$	0.969697	0.989899
$C_A(2000) \geq CounterC_A(2000)$	0	0
$CounterC_A(2000) \geq C_A(2020)$	0.5959596	0.1111111
$C_A(2020) \geq CounterC_A(2000)$	0.8585859	0
$CounterC_A(2020) \geq C_A(2020)$	0	0
$C_A(2020) \geq CounterC_A(2020)$	0.969697	1
$CounterC_A(2020) \geq C_A(2000)$	0.8383838	0
$C_A(2000) \geq CounterC_A(2020)$	0.5959596	0.1111111

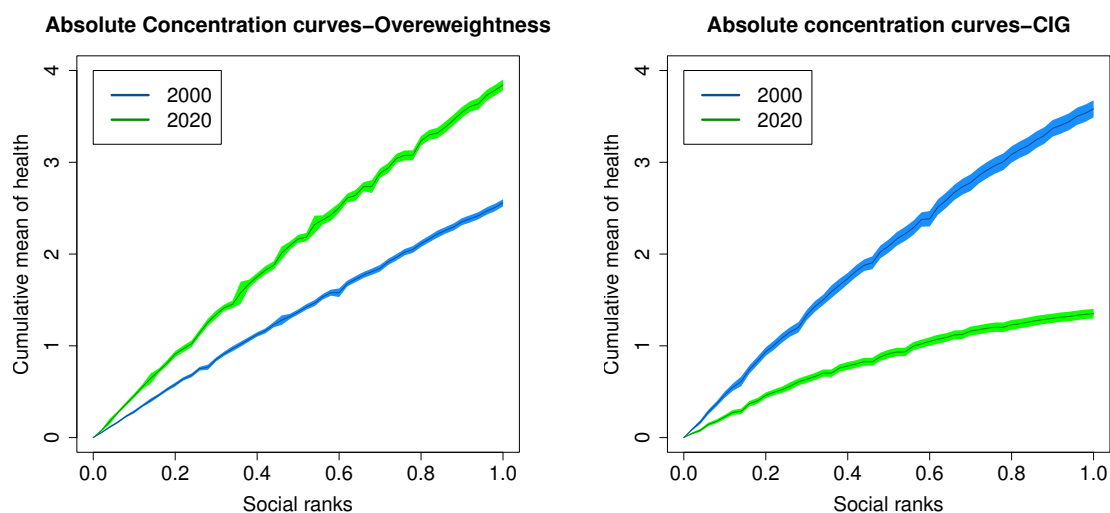
Note: $CounterC_A(2020)$ and $CounterC_A(2000)$ refer to the counterfactual of the absolute concentration curve of 2020 with return of 2000 and the counterfactual of the absolute concentration curve of 2000 with return of 2020, respectively.

FIGURE 1.1: Relative concentration curves



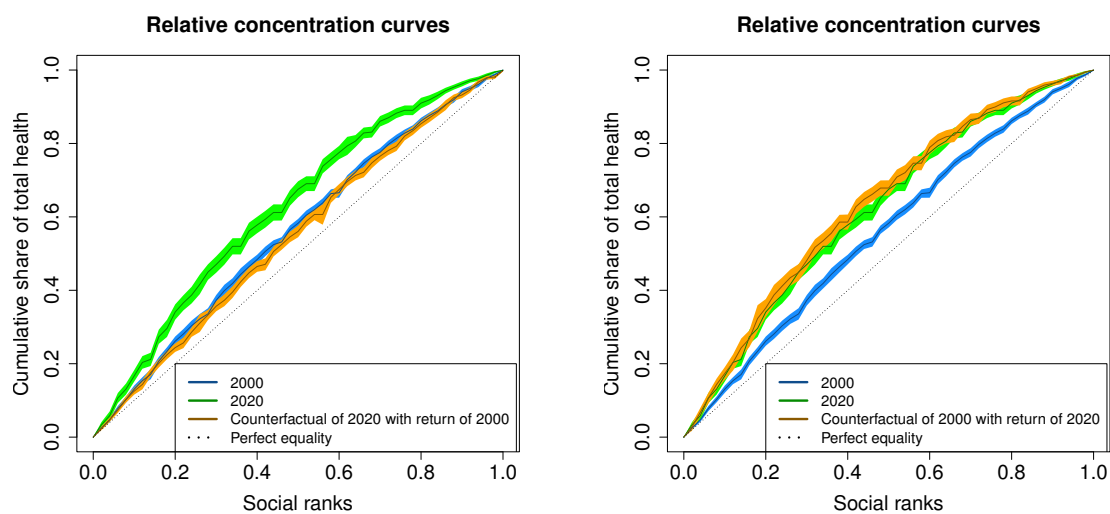
Note: The figure plots the relative concentration curves of overweightness in the left and cigarette consumption in the right for 2000 and 2020.

FIGURE 1.2: Absolute concentration curves



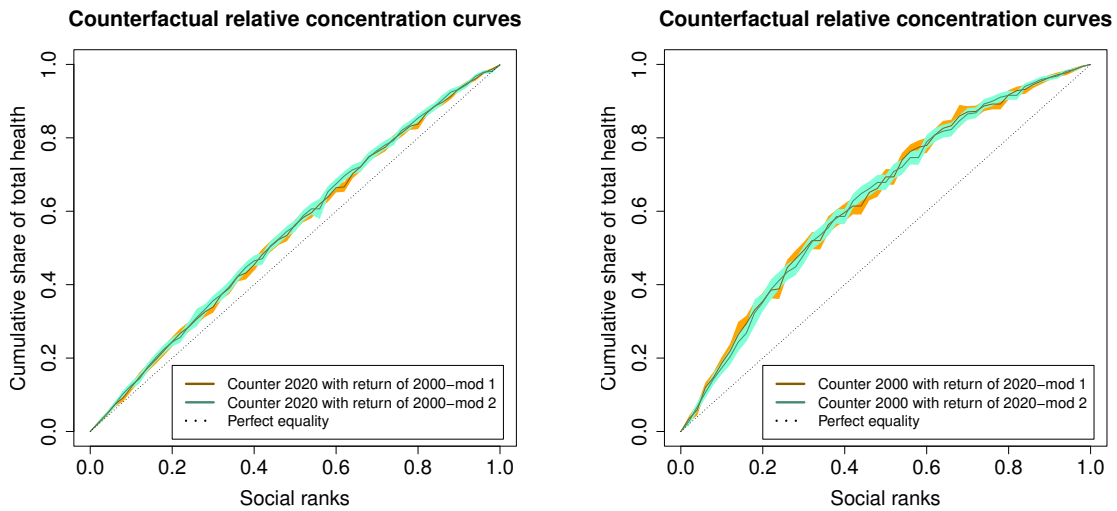
Note: The figure plots the absolute concentration curves of overweightness in the left and cigarette consumption in the right for 2000 and 2020.

FIGURE 1.3: Counterfactual relative concentration curves- Cigarette consumption (Model 2)



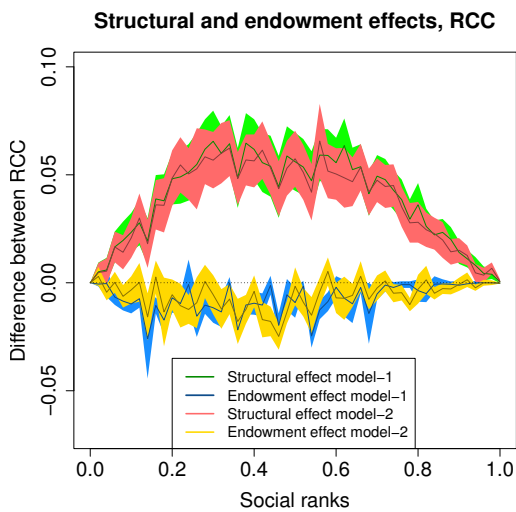
Note: The figure plots the relative concentration curves of cigarette consumption for 2000 and 2020 and their counterfactual curves estimated using the reweighting approach for Model 2 defined in Subsection 1.3.2. The counterfactual curve designates the curve that would have prevailed if individual characteristics had remained as in year 1, while individuals had been rewarded according to the return system of year 2.

FIGURE 1.4: Counterfactual relative concentration curves-Cigarette consumption (Two models)



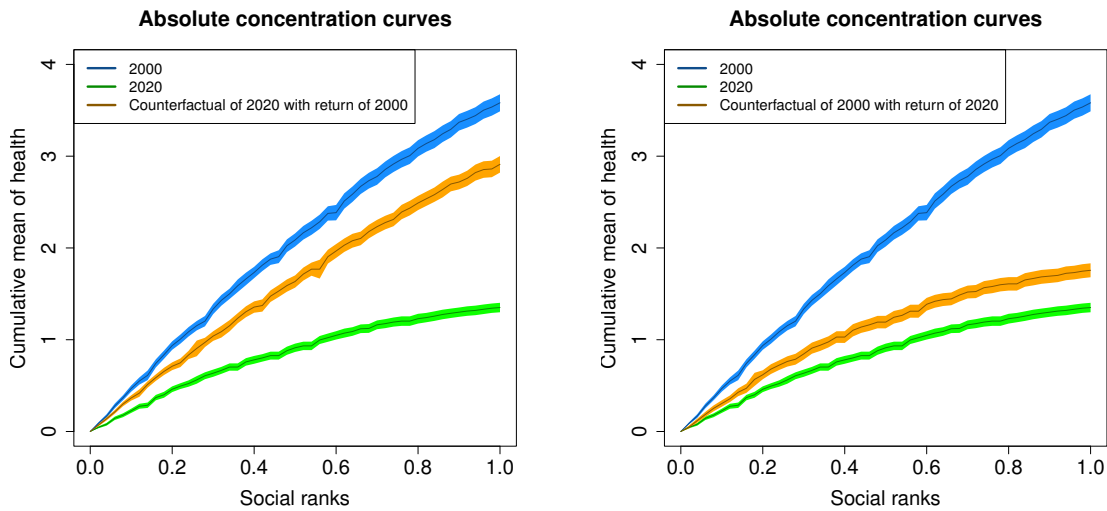
Note: The figure compares the counterfactual relative concentration curves of cigarette consumption in 2000 and 2020, estimated using the reweighting approach including the two models defined in Subsection 1.3.2.

FIGURE 1.5: Endowment and structural effects curves-Relative concentration curves (RCC) of cigarette consumption



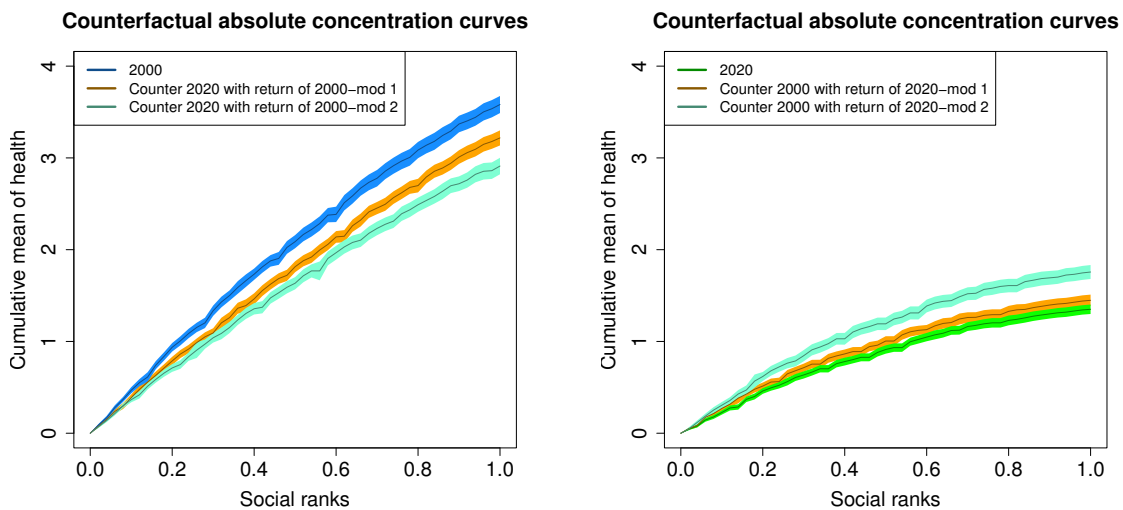
Note: The figure plots the endowment and the structural effect estimated using the reweighting approach to decompose the change in the relative concentration curves of cigarette consumption between 2020 and 2000, and compares the two models defined in Subsection 1.3.2.

FIGURE 1.6: Counterfactual absolute concentration curves-Cigarette consumption (Model 2)



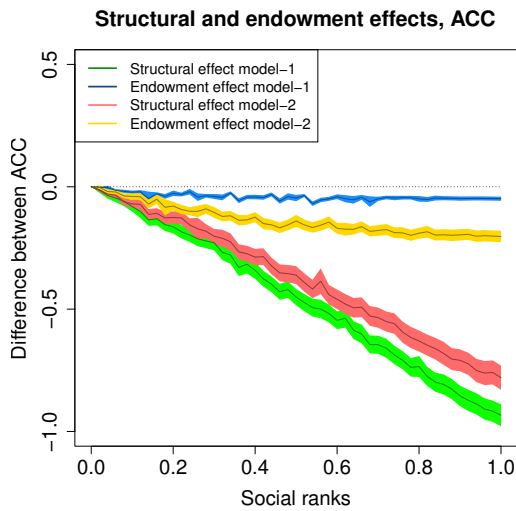
Note: The figure plots the absolute concentration curves of cigarette consumption for 2000 and 2020, along with their counterfactual curves estimated using the reweighting approach for Model 2 defined in Subsection 1.3.2. The counterfactual curve designates the curve that would have prevailed if individual characteristics had remained as in year 1, while individuals had been rewarded according to the return system of year 2.

FIGURE 1.7: Counterfactual absolute concentration curves-Cigarette consumption (Two models)



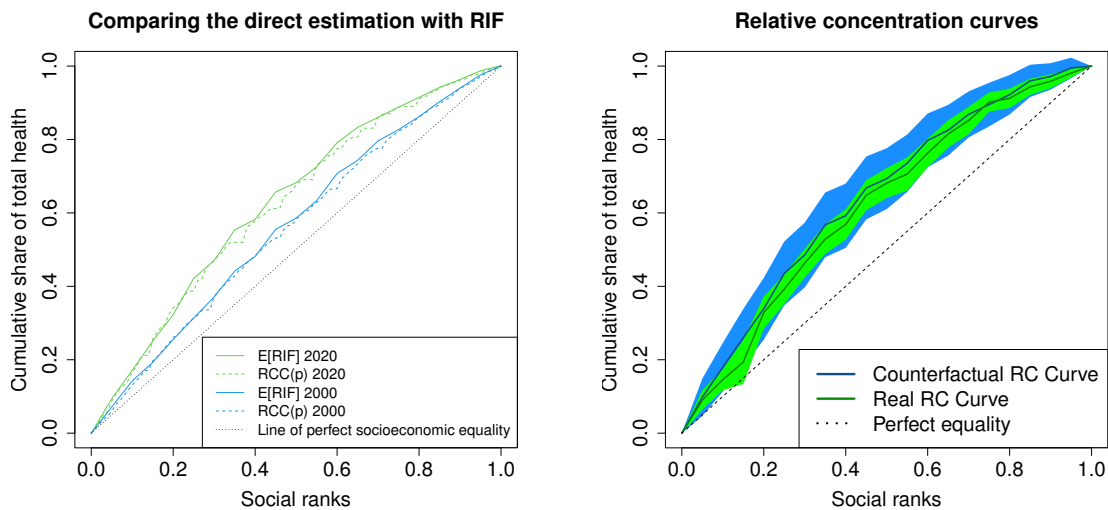
Note: The figure compares the counterfactual absolute concentration curves of cigarette consumption in 2000 and 2020, estimated using the reweighting approach including the two models defined in Subsection 1.3.2.

FIGURE 1.8: Endowment and structural effects curves-Absolute concentration curves (ACC) of cigarette consumption



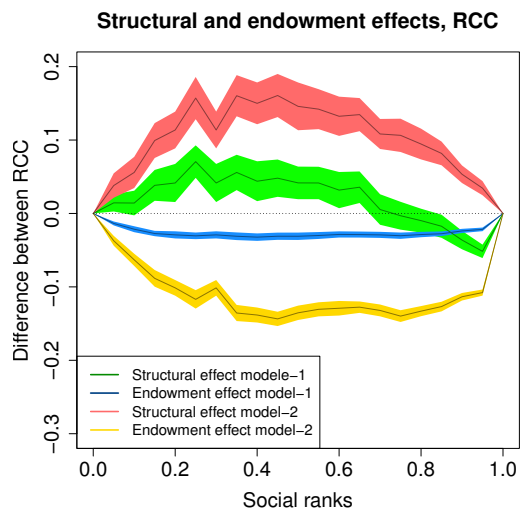
Note: The figure plots the endowment and the structural effect estimated using the reweighting approach to decompose the change in the absolute concentration curves of cigarette consumption between 2020 and 2000, and compares the two models defined in Subsection 1.3.2.

FIGURE 1.9: RIF model fit of the relative concentration curves-Cigarette consumption



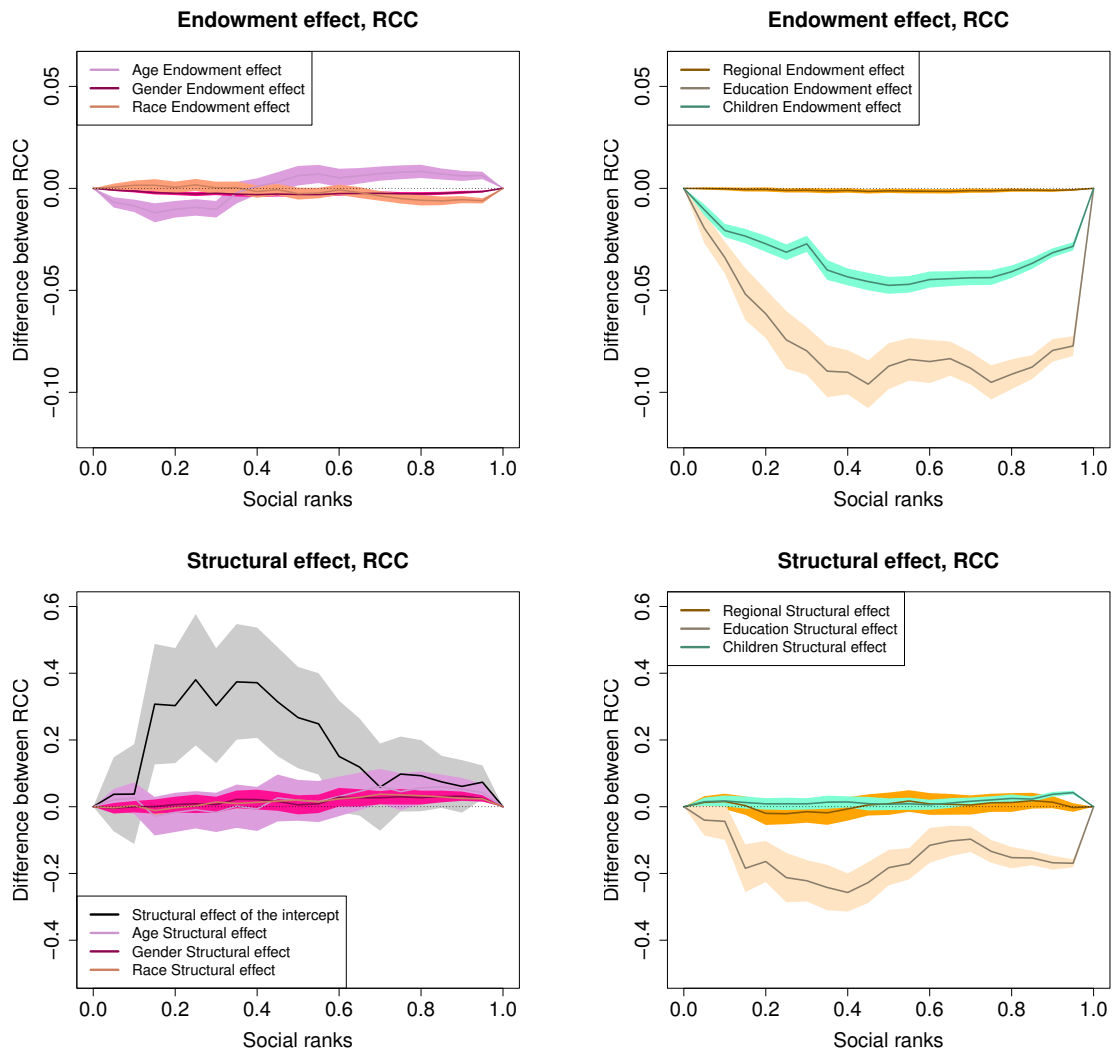
Note: In the left panel, the figure compares the expectation of the recentred influence function (RIF) of the relative concentration curves of cigarette consumption with the observed curves of 2000 and 2020. In the right panel, the figure compares the observed relative concentration curve of 2020 with the counterfactual curve estimated using an out-of-sample method and the RIF regression as explained in Subsection 1.3.2.

FIGURE 1.10: Structural and endowment effects-Relative concentration curves of cigarette consumption



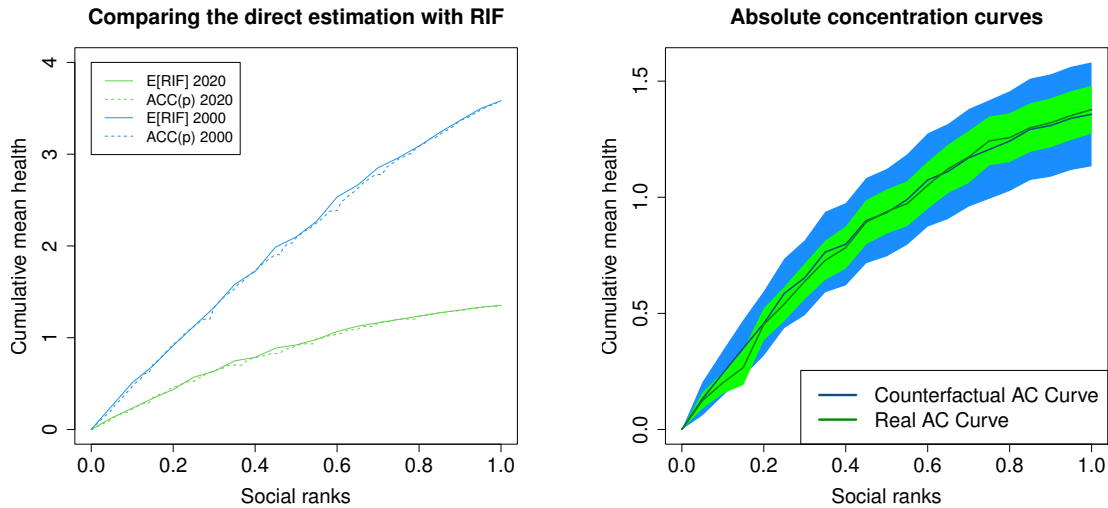
Note: The figure plots the endowment and the structural effect estimated using the RIF regression approach to decompose the change in the relative concentration curves of cigarette consumption between 2020 and 2000, and compares the two models defined in Subsection 1.3.2.

FIGURE 1.11: Contribution of each variable to the structural and the endowment effects in Model 2-Relative concentration curve of cigarette consumption



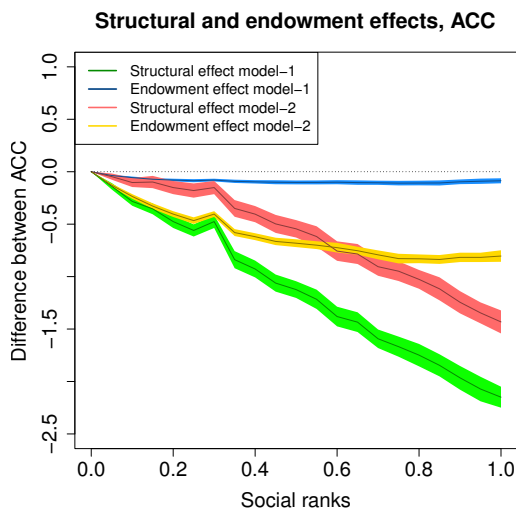
Note: The figure plots the contributions of each variable to the endowment (top panels) and the structural effect (bottom panels) estimated using the RIF regression approach of model 2 defined in Subsection 1.3.2 to decompose the change in the relative concentration curves of cigarette consumption between 2020 and 2000.

FIGURE 1.12: RIF model fit of the absolute concentration curves-
Cigarette consumption



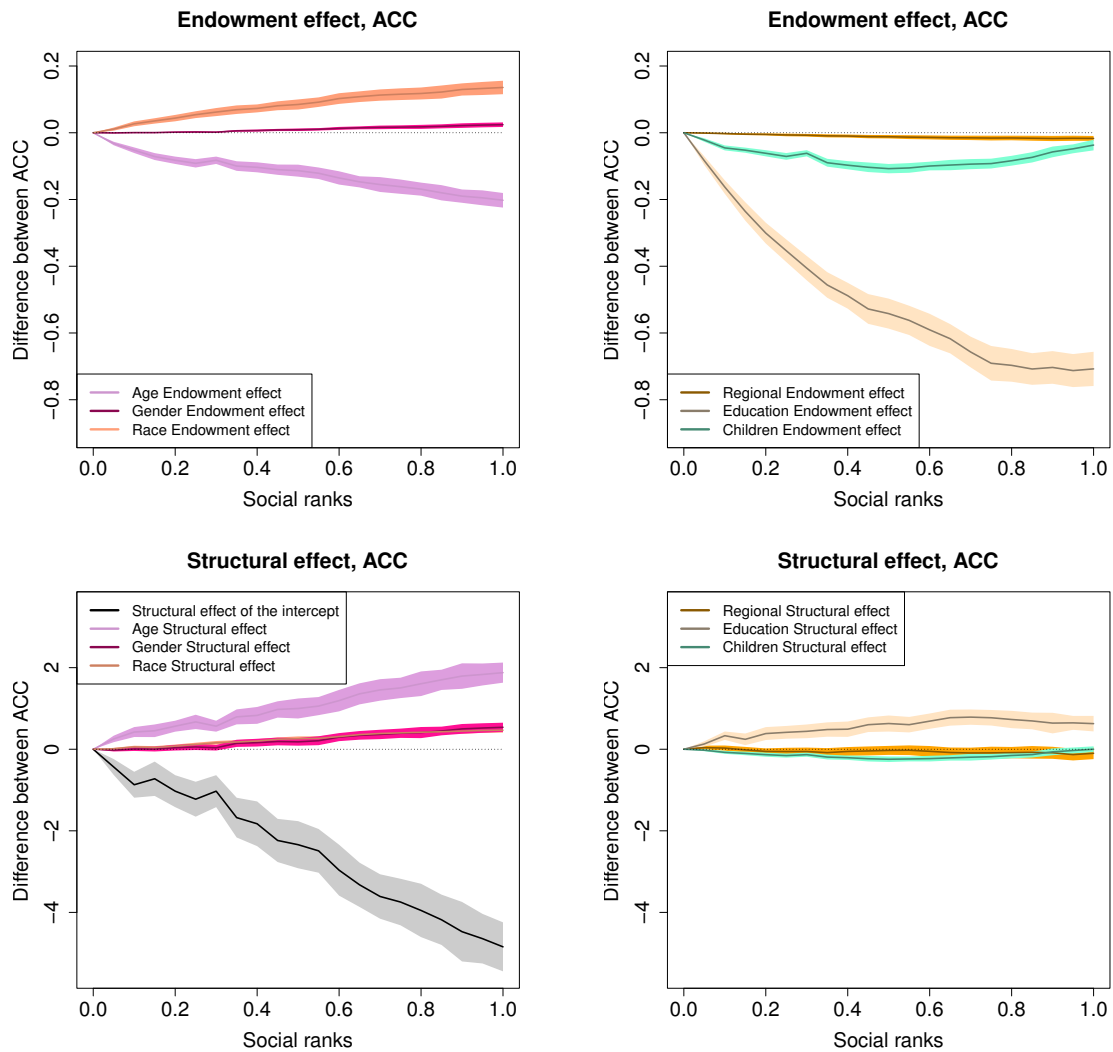
Note: In the left panel, the figure compares the expectation of the recentred influence function (RIF) of the absolute concentration curves of cigarette consumption with the observed curves of 2000 and 2020. In the right panel, the figure compares the observed absolute concentration curve of 2020 with the counterfactual curve estimated using an out-of-sample method and the RIF regression as explained in Subsection 1.3.2.

FIGURE 1.13: Structural and endowment effects-Absolute concentra-
tion curves of cigarette consumption



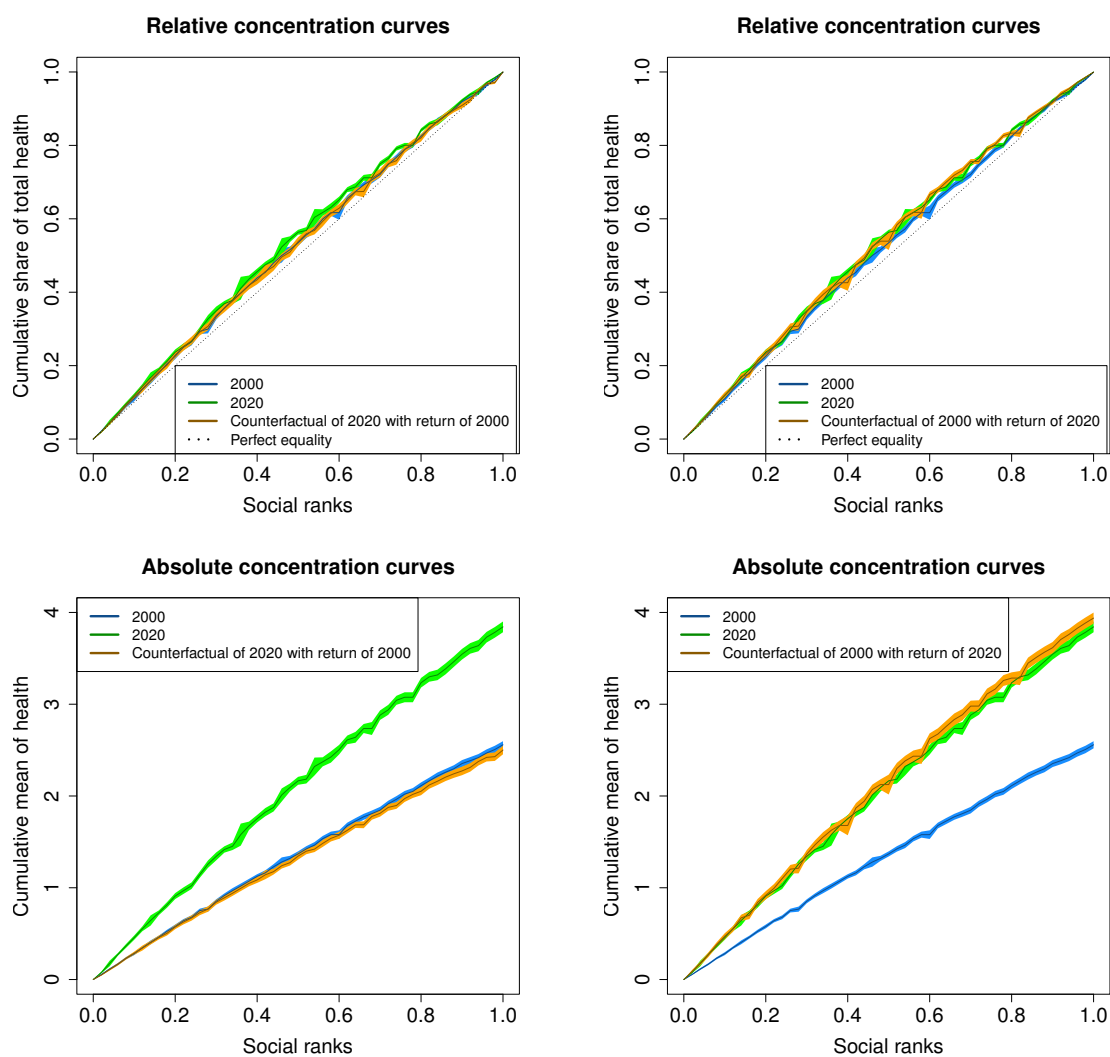
Note: The figure plots the endowment and the structural effect estimated using the RIF regression approach to decompose the change in the absolute concentration curves of cigarette consumption between 2020 and 2000, and compares the two models defined in Subsection 1.3.2.

FIGURE 1.14: Contribution of each variable to the structural and the endowment effects in Model 2-Absolute concentration curve of cigarette consumption



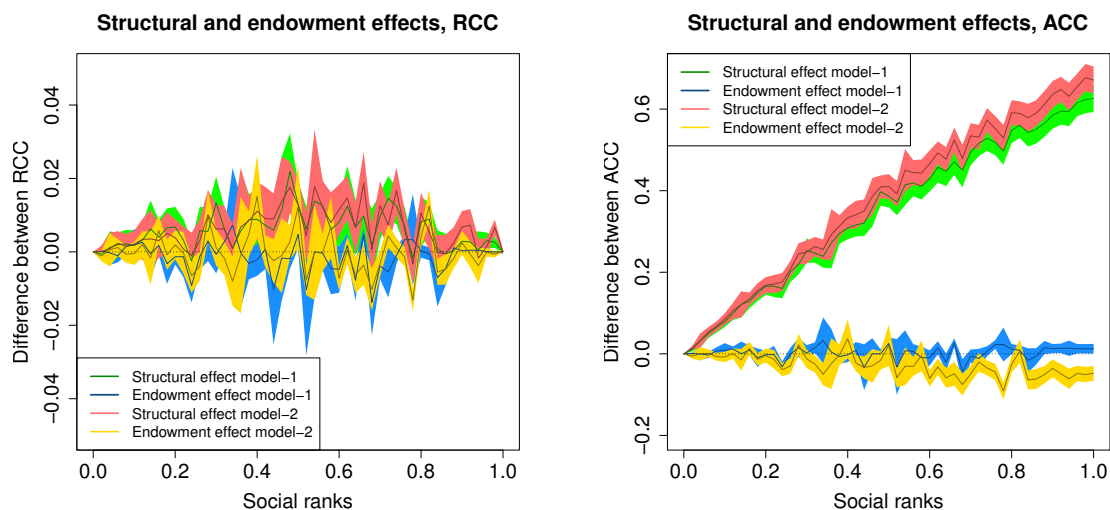
Note: The figure plots the contributions of each variable to the endowment (top panels) and the structural effect (bottom panels) estimated using the RIF regression approach of model 2 defined in Subsection 1.3.2 to decompose the change in the absolute concentration curves of cigarette consumption between 2020 and 2000.

FIGURE 1.15: Counterfactual curves-Overweightness (Model 2)



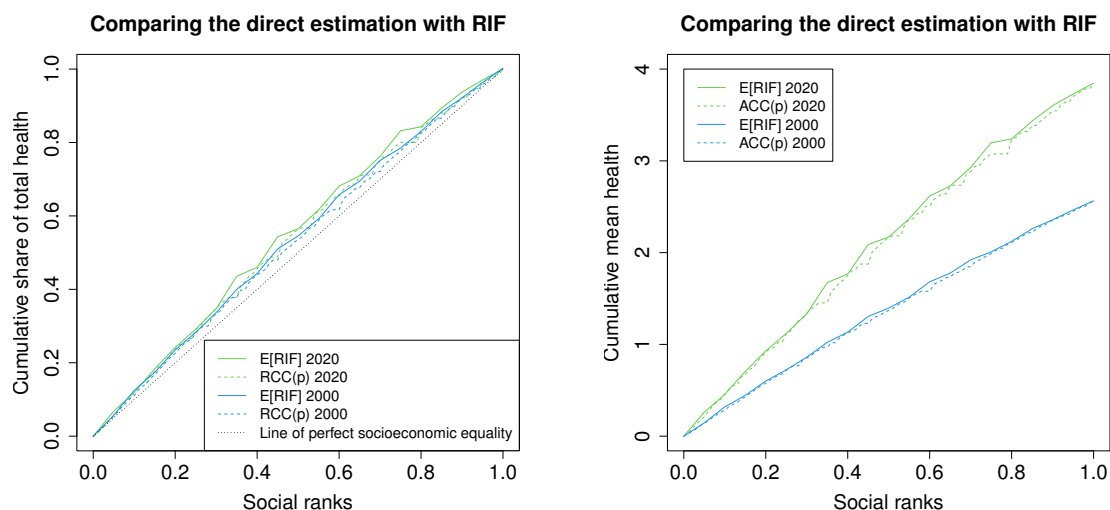
Note: The figure plots the relative (top panels) and absolute (bottom panels) concentration curves of overweightness for 2000 and 2020, along with their counterfactual curves estimated using the reweighting approach for Model 2 defined in Subsection 1.3.2. The counterfactual curve designates the curve that would have prevailed if individual characteristics had remained as in year 1, while individuals had been rewarded according to the return system of year 2.

FIGURE 1.16: Endowment and structural effects curves-Overweightness



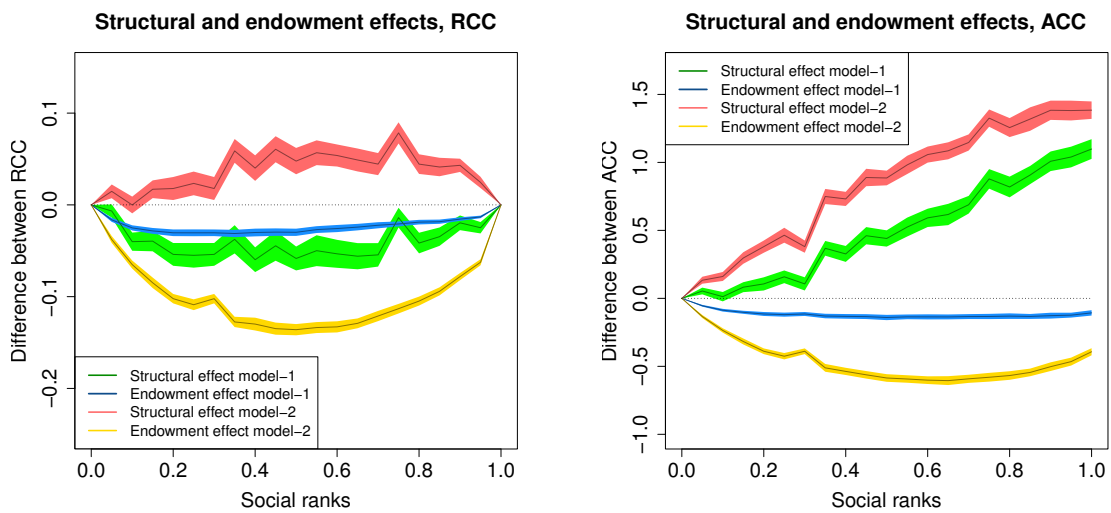
Note: The figure plots the endowment and the structural effect estimated using the reweighting approach to decompose the change in the relative (left panel) and the absolute (right panel) concentration curves of overweightness between 2020 and 2000. It compares the two models defined in Subsection 1.3.2.

FIGURE 1.17: RIF model fit of the relative and the absolute concentration curves-Overweightness



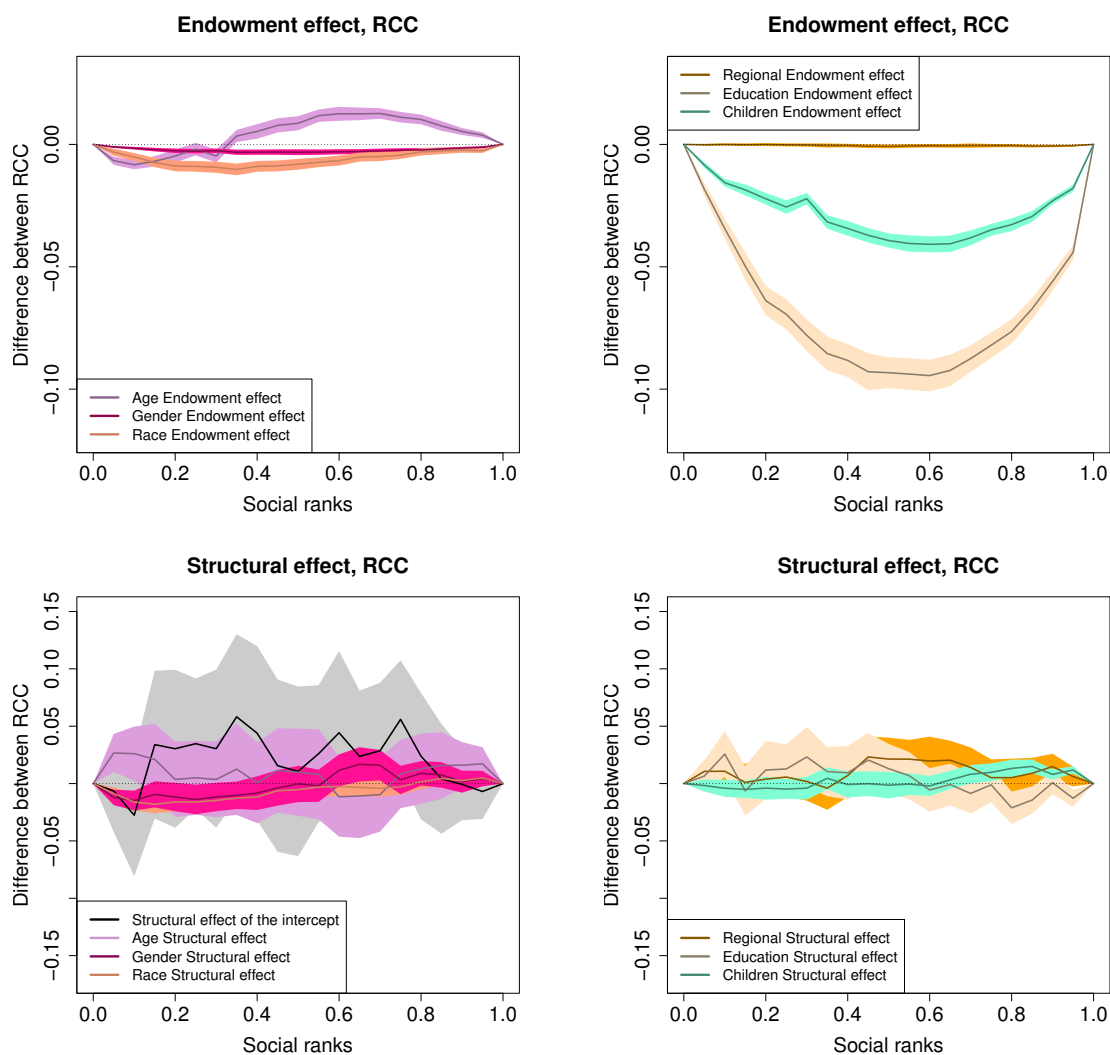
Note: The figure compares the expectation of the recentred influence function (RIF) of the relative (left panel) and the absolute (right panel) concentration curves of overweightness with the observed curves of 2000 and 2020.

FIGURE 1.18: Endowment and structural effects curves



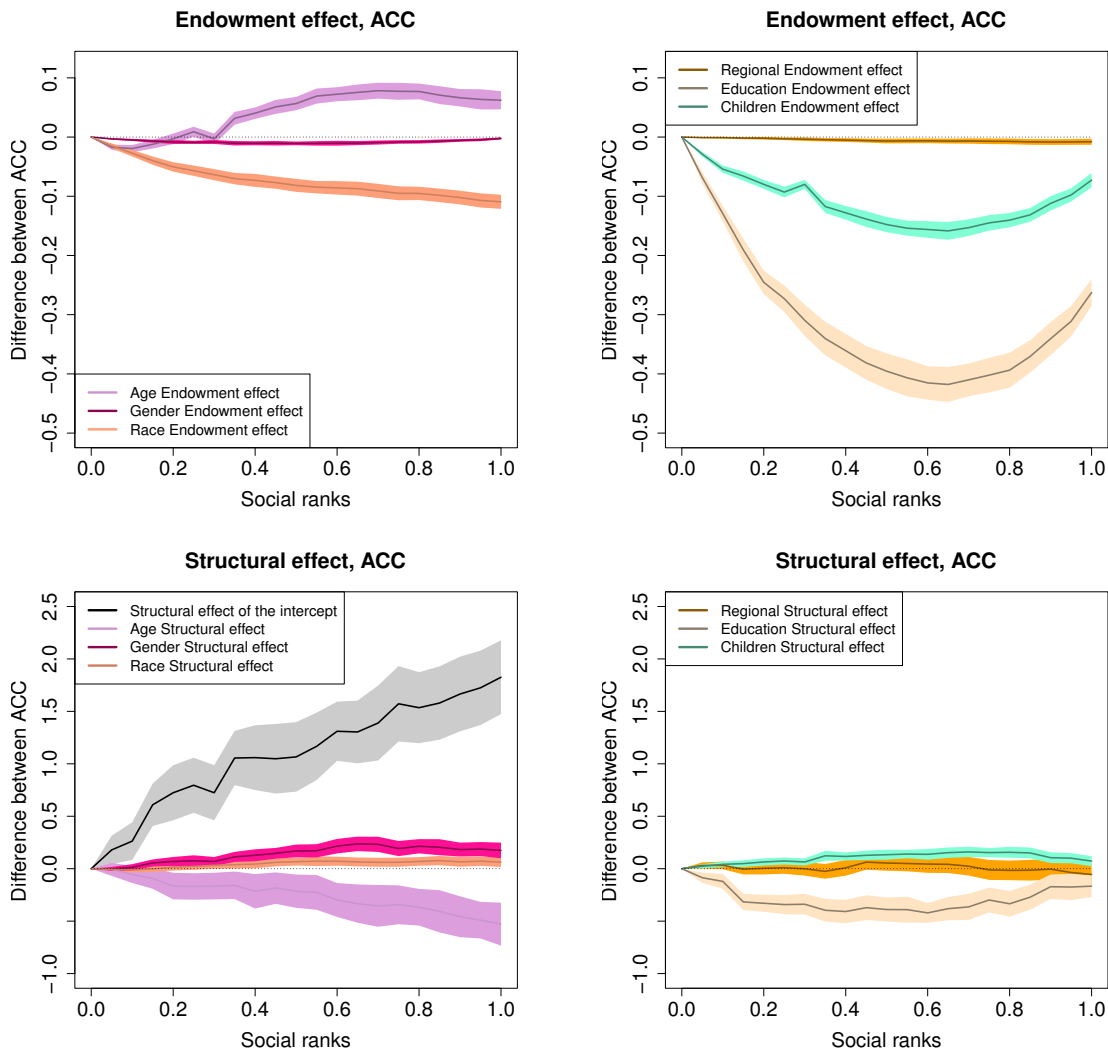
Note: The figure plots the endowment and the structural effect estimated using the RIF regression approach to decompose the change in the relative (left panel) and the absolute (right panel) concentration curves of overweightness between 2020 and 2000. It compares the two models defined in Subsection 1.3.2.

FIGURE 1.19: Contribution of each variable to the structural and the endowment effects in Model 2-Relative concentration curve of overweightness



Note: The figure plots the contributions of each variable to the endowment (top panels) and the structural effect (bottom panels) estimated using the RIF regression approach of model 2 defined in Subsection 1.3.2 to decompose the change in the relative concentration curves of overweightness between 2020 and 2000.

FIGURE 1.20: Contribution of each variable to the structural and the endowment effects in Model 2-Absolute concentration curve of overweightness



Note: The figure plots the contributions of each variable to the endowment (top panels) and the structural effect (bottom panels) estimated using the RIF regression approach of model 2 defined in Subsection 1.3.2 to decompose the change in the absolute concentration curves of overweightness between 2020 and 2000.

Chapter 2

The Cognitive Health of Black Americans: Assessing the Impact of De Facto versus De Jure Segregation

2.1 Introduction

Cognitive health is a critical dimension of well-being in older adulthood, underpinning independence, quality of life, and longevity. With an aging population and the growing burden of dementia in the United States, understanding the determinants of cognitive decline has become a pressing public health concern (CDC and Alzheimer's Association, 2007). While aging remains the primary risk factor, research consistently reveals stark racial disparities: older Black Americans are nearly twice as likely as their White counterparts to develop Alzheimer's disease or related dementias (Manly et al., 2022; Power et al., 2021). These disparities cannot be fully explained by genetic predisposition (Yaffe et al., 2013) or individual health behaviors, pointing instead to broader social and historical factors (Bailey et al., 2017; Williams and Collins, 2001).

Among the most consequential historical factors is the legacy of Jim Crow laws, which enforced de jure racial segregation across the Southern United States until the mid-1960s. These laws institutionalized racial inequality by limiting access to quality education, healthcare, housing, and employment opportunities for Black Americans,

with potential consequences for health trajectories over the life course. In contrast, while states outside the South did not codify segregation in law, they often maintained racial separation through informal mechanisms—so-called de facto segregation—such as discriminatory housing markets, social norms, and institutional practices. While a growing literature has examined the health implications of contemporary de facto segregation, little is known about the long-term health consequences of legalized segregation.

This paper aims to fill that gap by examining whether non-exposure to de jure segregation—despite the continued presence of de facto segregation—was associated with improved cognitive health outcomes later in life. We treat individuals born outside the Jim Crow South as the treatment group, representing early-life non-exposure to legalized segregation, and those born within the South as the control group, exposed to de jure segregation.

Rather than relying solely on average treatment effects, we adopt a distributional approach that captures how these effects vary across the income distribution. More specifically, we combine a Difference-in-Differences (DiD) strategy with a Recentered Influence Function (RIF) regression to estimate the impact of the absence of Jim Crow laws on both health achievement (via absolute concentration curves) and socioeconomic health inequality (via relative concentration curves). This methodological innovation enables us to evaluate not only whether cognitive outcomes improved, but also who benefited, and whether these gains helped reduce socioeconomic health inequality.

Our findings show that the benefits of non-exposure to Jim Crow laws were not uniform. Cognitive improvements were concentrated among individuals up to the 85th percentile of income, with little effect observed at the top of the distribution, leading to no statistically significant effect on average. These findings underscore the limitations of relying solely on average treatment effects to understand the distributional consequences of structural policy change.

In sum, this paper offers new evidence on the persistent effects of early-life exposure to legalized segregation on health in older age. It highlights the importance of examining

not only average policy effects but also their distribution across socioeconomic groups. By uncovering cognitive health improvements that are concentrated below the top of the income distribution, this paper demonstrates that the consequences of dismantling segregationist policies extend beyond averages to the structure of inequality itself. In doing so, it contributes to three key literatures: (1) the long-term health consequences of historical segregation; (2) the economics of health inequality; and (3) applied policy evaluation using distributional methods. To our knowledge, this is the first study to estimate the causal and distributional impact of Jim Crow laws on cognitive health in later life, offering new insight into how the legacy of segregation continues to shape health disparities decades later with a distributional lens.

The remainder of the paper is structured as follows. Section 2.2 provides an overview of Jim Crow laws and a review of the literature on segregation and health. Section 2.3 outlines the theoretical framework used in our analysis. Section 2.4 presents the dataset and descriptive statistics, offering an empirical foundation for our study. In Section 2.5, we describe our empirical strategy, detailing the methodological choices and estimation techniques employed. Section 2.6 reports our findings, highlighting the impact of Jim Crow laws on cognitive health across socioeconomic ranks. Finally, Section 2.8 presents concluding remarks, discussing the broader implications of our results and potential avenues for future research.

2.2 Background and literature review

2.2.1 Jim Crow laws in the South

Segregation, the systematic separation of people based on race, ethnicity, or socioeconomic status, is deeply rooted in the history of the United States, originating from the slavery era and later institutionalized by Jim Crow laws.

Jim Crow laws, which were state and local laws implemented across the southern United States from the 1870s until the mid-1960s, enforced racial segregation in public facilities, including schools, transportation, parks, restrooms, and restaurants, among others. Figure 2.1 shows the 17 states and the District of Columbia where Jim Crow

laws were implemented.¹ The laws were named after Thomas Dartmouth Rice's minstrel show character in the 19th century, depicting an African American in a derogatory manner. After the Civil War, the South started passing laws that oppressed newly freed African Americans through segregation, starting with the Separate Car Act of 1890 and upheld by the Supreme Court decision in *Plessy v. Ferguson* (1896), which established the "separate but equal" doctrine, allowing for racially segregated facilities as long as they offer the same quality of services. In reality, the facilities provided to African Americans were vastly inferior in quality and underfunded (Klarman, 2007).

The dismantling of Jim Crow laws began during the Civil Rights Movement of the 1950s and 1960s. The Supreme Court's landmark decision in *Brown v. Board of Education* (1954), declaring racial segregation in public schools unconstitutional, marked the beginning of the end for Jim Crow laws. However, the decision faced significant resistance, and many southern states delayed or avoided implementing desegregation. It was not until 1964 that the passage of the Civil Rights Act effectively ended legal segregation in public places. It outlawed discrimination based on race, color, religion, sex, or national origin. It gave the federal government the authority to enforce the law of school desegregation, which was lacking in the *Brown v. Board of Education* decision. Subsequent legislation, including the Voting Rights Act of 1965 and the Fair Housing Act of 1968, further dismantled legalized segregation.

2.2.2 Literature review

Although the legal aspect of segregation was abolished years ago, the legacy of segregation continues to affect African Americans and other minority groups, shaping the social landscape of the U.S. Segregation has been shown to have long-lasting effects on various aspects of life, including education, the economic sphere, and, not to mention, health. In fact, a higher level of poverty usually depicts segregated areas, a poorer housing environment, and poorer educational and economic opportunities, all of which negatively affect health outcomes (Williams, Lawrence, and Davis, 2019).

¹Limited laws were adopted in Wyoming, Kansas, New Mexico, and Arizona, such as banning intermarriage or requiring separate rooms in schools.

A growing body of research highlights the profound impact of social and environmental determinants on health and well-being. Racial segregation has been identified as a fundamental cause of health disparities among Black populations (Bailey et al., 2017; Williams and Collins, 2001).

Numerous studies have examined contemporary patterns of residential segregation, typically quantified using measures such as the dissimilarity index or the isolation index (Massey and Denton, 1988). They all agreed on the negative association between black segregation and poor health outcomes (Kramer and Hogue, 2009), including weak birth outcomes and high rate of stillbirth (Austin, Harper, and Strumpf, 2016; Vu, Green, and Swan, 2024; Williams et al., 2018), obesity and adverse health behaviors (Piontak and Schulman, 2016; Walsemann and Bell, 2010), and asthma (Alexander et al., 2017). Similar findings have been reported regarding adult and older health (Robert and Ruel, 2006; Sudano et al., 2013), with segregation linked to poorer self-rated health (Anderson, 2016; Do, Frank, and Iceland, 2017), obesity and adverse health behaviors (Corral et al., 2012), and mental health issues (Reddy and Brown, 2022).

Despite this extensive research, only a few studies have specifically explored the impact of Jim Crow laws on health, and even fewer have investigated their effects on cognitive health. For example, Krieger et al. (2013) found that the Black infant death rate was 1.19 times higher in Jim Crow states compared to non-Jim Crow states between 1960 and 1964. Additionally, Krieger, Jahn, and Waterman (2017) demonstrated that being born in a Jim Crow state was a predictor of breast cancer more than 50 years later. While Walsemann et al. (2023) highlighted the positive effect of desegregated schools on the cognitive functioning of Black adults exploiting the abolition of Jim Crow laws, other studies (Caunca et al., 2020; Meyer et al., 2021) settled for using residential segregation. Though Lin et al. (2024) used attending segregated school before college as a determinant of negative cognitive health outcomes, it remains different from the effect of Jim Crow laws.

Overall, the existing research documents the detrimental effects of segregation on health, utilizing various forms of segregation such as residential segregation, school (de)segregation, and Jim Crow laws. However, few studies have assessed the causal

impact of segregation on health outcomes (Austin, Harper, and Strumpf, 2016; Vu, Green, and Swan, 2024), and, to our knowledge, none have specifically examined the causal effect of de jure segregation under Jim Crow laws on cognitive health.

The legacy of Jim Crow laws, while officially abolished decades ago, continues to cast a long shadow over the health and cognitive functioning of African Americans. Given the strong evidence linking early-life experiences and human capital investment to later-life cognitive outcomes (Currie and Almond, 2011; Heckman, Pinto, and Savelyev, 2013), and the well-documented predictive role of birth endowments for both adult outcomes and intergenerational effects (Currie, 2011), we analyse how Jim Crow laws may affect cognitive decline among Black Americans. These laws restricted access to quality education, healthcare, and public facilities, setting the stage for disparities in cognitive health that persist into old age. In this paper, we seek to assess the causal effect of de facto (informal) segregation, compared to legalized segregation under Jim Crow laws, on the cognitive functioning of Black older adults in the U.S.

2.3 Theoretical framework

To assess the causal effect of Jim Crow laws on cognitive health, a commonly used approach is to look at the average treatment effect. In this case, a natural approach to consider would be the DiD strategy. In this context, by comparing changes in outcomes for Black and Non-Black individuals in the treated and control groups, DiD effectively isolates the average treatment effect. While DiD is widely used, it overlooks the heterogeneity in the laws' impact across social strata, focusing solely on the average treatment effect. Examining the effect across ranks is essential to understanding how policies or interventions affect different segments of the population and identifying who benefits the most or least. This ensures that policies are better targeted to address equity and distributional concerns. For instance, a policy may yield a significant average effect. Still, its benefits could be concentrated among individuals at the upper end of the distribution, leaving those at the lower end with little to no advantage. Furthermore, by analyzing effects across socioeconomic ranks, researchers can uncover whether interventions reduce or exacerbate inequalities, capture nuanced patterns that

averages may hide, and design policies that better support vulnerable groups and guide policymakers in allocating resources more effectively.

2.3.1 Socioeconomic health inequality measures

To address this limitation, we propose an alternative approach that moves beyond averages to capture the treatment effect across the entire distribution of socioeconomic ranks, while still encompassing the average effect as a special case. The proposed method allows us to conduct our analysis using canonical measures of health inequality: the absolute and relative concentration curves. The relative concentration curve is a bivariate functional of the joint distribution of health and income (or any other measure of socioeconomic status). It provides a comprehensive view of how health is distributed among individuals, from the most disadvantaged to the least disadvantaged. Formally, the relative concentration curve plots the cumulative proportion of health H against the cumulative proportion of the population ranked by income Y from the most disadvantaged to the least disadvantaged.

While the relative concentration curve primarily measures health inequalities, Wagstaff (2002) highlighted a crucial concern: policymakers may be interested not only in health inequalities but also in average health outcomes. For example, if wealthier individuals experience a decline in health over time and converge to the same health level as poorer individuals, health inequality would decrease, but average health would also deteriorate. To address this concern, Wagstaff introduced the concept of “achievement” in health, which captures both the average level of health and the socioeconomic inequality in its distribution in a single measure. Specifically, the absolute concentration curve is derived by multiplying the relative concentration curve by the population’s average health. Intuitively, at each social rank q , it represents the cumulative contribution of individuals ranked q or lower to the overall average health. Consequently, it provides a graphical representation of health achievement (good health) or health shortfall (ill health) by socioeconomic ranks, which is more relevant from a policy perspective.

Assume that H , health, and Y , income have a joint distribution F_{HY} . The absolute concentration curve (C_A) and the relative concentration curve (C_R) are formally

defined as

$$C_A(q, F_{HY}) = \int_0^q h(u) du$$

$$C_R(q, F_{HY}) = \frac{1}{\mu_H} \int_0^q h(u) du$$

where

$$h(q) = E[H|Y = F_Y^{-1}(q)]$$

and

$$\mu_H = \int_0^1 h(u) du$$

Figure 2.2 provides a graphical representation of the relative concentration curve alongside the 45-degree line, which represents perfect relative socioeconomic health equality. The further the curve deviates from the 45-degree line, the more pronounced the health inequalities become. Additionally, the curve's position relative to the 45-degree line is informative. Specifically, if the curves lie below, the inequalities are pro-rich for good health and pro-poor for ill-health, and vice versa if the curves lie above. Figure 2.3 illustrates the absolute concentration curve. At rank $q = 1$, this curve represents the population's average health. Since the absolute concentration curve incorporates inequality, its shape provides insights regarding socioeconomic health inequalities. A convex curve indicates that health inequalities are pro-rich for good health and pro-poor for ill-health, whereas a concave curve suggests the opposite pattern.

Moreover, these curves enable the identification of robust rankings of joint distributions of health and income that remain valid regardless of the specific health achievement/shortfall or socioeconomic health inequality index employed (Khaled, Makdissi, and Yazbeck, 2018; Makdissi and Yazbeck, 2014). Although the absolute and relative concentration curves offer a comprehensive and robust representation of health achievement/shortfall and socioeconomic health inequality, some researchers advocate using specific indices, such as the concentration index or the achievement/shortfall index.²

²The concentration index is equal to twice the area between the relative concentration curve and the 45-degree line.

The achievement/shortfall index equals the average health multiplied by 1 minus the concentration index.

2.3.2 Identification

Our primary interest is to assess the treatment effect of de facto versus de jure segregation under Jim Crow laws on the cognitive health of older adults using a beyond-average-treatment-effect approach. Following the causal inference framework, identifying this effect requires a conceptual framework based on a counterfactual setting. In the context of this paper, this means that we need to construct a counterfactual concentration curve that represents what would have occurred in the absence of the treatment. Under the assumption of selection on observables, a reweighting approach using a propensity score would be sufficient to construct these counterfactual concentration curves and estimate the causal effect. However, when selection is driven by unobservable characteristics—which is generally the case—reweighting alone is insufficient to identify the causal treatment effect.

While the causal inference literature offers a wide range of tools, they are not readily applicable to our research question. Specifically, given that our object of interest is a bivariate functional of the joint distribution of health and income rather than the mean of the outcome variable, standard econometric tools cannot be directly applied to assess the impact of a marginal change in an explanatory variable on the distributional statistic.

To address this issue, we employ the framework developed by Bchi, Makdissi, and Yazbeck (2024) to estimate the marginal effect of small shifts in the distribution of explanatory variables on a bivariate functional of the joint distribution of health and income. Bchi, Makdissi, and Yazbeck (2024) exploit and extend the recentred influence function (RIF) approach proposed by Firpo, Fortin, and Lemieux (2009), which offers a bridge between the inherently nonlinear distributional statistics and the standard regression framework.

RIF approach of the concentration curves

The RIF represents the two leading terms of the von Mises (1947) first-order linear approximation of the distributional statistic, leveraging a key property: the expectation of the RIF is equal to the statistic itself. Consequently, in the presence of covariates X , the law of iterated expectations can be applied to assess unconditional expectations

$E[RIF]$. Specifically, the unconditional expectation of the RIF can be decomposed into its conditional expectation given X —that is, $E[RIF] = E[E[RIF | X]]$. This conditional expectation, $E[RIF | X]$, can then be readily estimated using regression methods, allowing us to evaluate how changes in the distribution of covariates influence the concentration curves.

The RIF of the absolute concentration curve, $RIF(h_i, y_i, C_A(q, F_{HY}))$, and the RIF of the relative concentration curve, $RIF(h_i, y_i, C_R(q, F_{HY}))$ provide linear approximations of $C_A(q, F_{HY})$ and $C_R(q, F_{HY})$, respectively. Hence, the impact of changes in the distribution of exogenous variables on the RIF serves as a first-order approximation of their impact on the concentration curves.

Bchi, Makdissi, and Yazbeck (2024) derived the expressions for the RIF of the absolute and relative concentration curves as follows:

$$\begin{aligned} RIF(h_i, y_i, C_A(q, F_{HY})) &= [q - \mathbb{1}(y_i \leq F_Y^{-1}(q))] \cdot E[H | Y = F_Y^{-1}(q)] \\ &\quad + \mathbb{1}(y_i \leq F_Y^{-1}(q)) \cdot h_i. \end{aligned} \quad (2.1)$$

$$\begin{aligned} RIF(h_i, y_i, C_R(q, F_{HY})) &= C_R(q, F_{HY}) \cdot \left[1 - \frac{h_i}{\mu_H}\right] + [q - \mathbb{1}(y_i \leq F_Y^{-1}(q))] \cdot \frac{E[H | Y = F_Y^{-1}(q)]}{\mu_H} \\ &\quad + \mathbb{1}(y_i \leq F_Y^{-1}(q)) \cdot \frac{h_i}{\mu_H}. \end{aligned} \quad (2.2)$$

Firpo, Fortin, and Lemieux (2009) demonstrated that the regression coefficients represent the marginal impact on the distributional statistic when the change in the distribution of covariates F_X is marginal. However, since the distribution of covariates is often different between the treatment and control groups, the assumption of marginal changes is not plausible. Therefore, for non-marginal changes in F_X , the linearization of the RIF may introduce non-negligible errors. Rothe (2015) highlighted the need to account for deviations from linearity in such cases. To address this issue, Firpo, Fortin, and Lemieux (2018) proposed using reweighting functions to adjust for substantial changes in the distribution. In this paper, we adapt this approach to our

empirical framework. Given that the choice of the reweighing approach depends on the empirical setting, the details are explained in section 2.5.

Causal distributional treatment effect model

As the RIF framework allows us to operate within a standard linear regression model, we can set up a counterfactual framework to evaluate the effects of de facto versus de jure segregation on both absolute and relative concentration curves. Specifically, for each rank q , we can estimate a DiD model with the RIFs for that rank on the left-hand side of the regression equation. This approach, referred to as **distributional DiD RIF regression**, naturally estimates the treatment effect as a curve since $RIF(h_i, y_i, C_A(q, F_{HY}))$ and $RIF(h_i, y_i, C_R(q, F_{HY}))$ are functions of the social rank q , which is continuous on $[0, 1]$. By estimating the treatment effect curves on absolute and relative concentration curves at each socioeconomic rank, this approach offers a more comprehensive understanding of how effects vary across the distribution and how non-exposure to Jim Crow laws has influenced health inequalities.

Moreover, the treatment curves provide a robust interpretation of the effect on any distributional statistic used, including the mean. More precisely, if the effect on the absolute concentration curve is significantly positive (negative) across all ranks, the treatment effect is positive (negative) on any health achievement/shortfall index used. Similarly, if the effect on the relative concentration curve is significantly positive (negative) across all ranks, the treatment effect is negative (positive) on any socioeconomic health inequality index used, including the canonical concentration index.

Furthermore, it is worth noting that at rank $q = 1$, the RIF of the absolute concentration curve simplifies to the health variable itself ($RIF(h_i, y_i, C_A(1, F_{HY})) = h_i$). In this case, the distributional DiD RIF regression becomes equivalent to a standard DiD regression with the health variable as the dependent variable, ensuring consistency with traditional methods for this specific rank.

2.4 Data and descriptive statistics

In this paper, we use public data from the 2016 wave of the Health and Retirement Study (HRS), a longitudinal, nationally representative survey of the U.S. population

aged 50 and older. Sponsored by the National Institute on Aging (grant number NIA U01AG009740) and conducted by the University of Michigan, the HRS covers a broad range of topics, including health, income, employment, and demographic characteristics. The majority of variables used in this study are sourced from the RAND HRS data—a longitudinal file containing cleaned, ready-to-use versions of the most frequently utilized HRS variables.³ Additionally, raw HRS files are used to derive variables not included in the RAND data, such as household members' income beyond that of respondents and their spouses.

2.4.1 Health outcome variables

While the diagnosis of cognitive decline, particularly dementia, typically requires neuropsychological tests administered by health professionals such as neurologists or psychiatrists, the cognitive tests used in large-scale, nationally representative surveys like the Health and Retirement Study (HRS) primarily assess cognitive functioning rather than provide a formal diagnosis. However, the Aging, Demographics, and Memory Study (ADAMS), a substudy of the HRS, conducted detailed neuropsychological and clinical assessments on a smaller subsample of the national dataset (Langa et al., 2005), alongside the cognitive tests administered in the HRS. Building on this work, Langa et al. (2009) developed a method to link the cognitive summary score derived from the HRS tests to dementia and mild cognitive decline (MCI) diagnoses in ADAMS. This method, known as the Langa–Weir classification, categorizes scores as follows: 0–6 indicates dementia, 7–11 indicates cognitive impairment without dementia, and 12 and above indicates normal cognitive decline. Furthermore, Crimmins et al. (2011) demonstrated that the HRS cognitive tests predict ADAMS diagnoses with 74% accuracy, underscoring their utility in large-scale population studies.

The health variables of interest are cognitive functioning indices. Cognition is assessed using multidimensional measures derived from a reduced version of the Telephone Interview for Cognitive Status (TICS), administered to new entrants or individuals aged 65 and older. These measures include:

³The RAND HRS Longitudinal File is a user-friendly dataset developed by RAND with funding from the National Institute on Aging and the Social Security Administration.

1. Immediate and delayed word recall: Memory is assessed by asking respondents to recall words from a list of 10 immediately and again after five minutes. The scores reflect the number of words recalled in each instance.
2. Serial 7's test: Working memory is evaluated by asking respondents to subtract seven from 100 for five consecutive trials. The score ranges from 0 to 5, based on the number of correct subtractions.
3. Counting backwards: Mental processing speed is assessed by asking respondents to count backwards continuously from 20 for 10 numbers. Two points are awarded for success on the first try, one on the second, and zero if unsuccessful.
4. Naming tasks: These include date naming (day, month, year, and day of the week, one point each), object naming (scissors and cactus, one point each), and naming the president and vice president (one point each).

Figure 2.4 illustrates the three summary indices derived from the cognitive tests. The word recall index summarizes performance on the immediate and delayed word recall tasks, with scores ranging from 0 to 20. The mental status index aggregates scores from the serial 7's test, counting, and naming tasks, ranging from 0 to 15. Finally, the total cognition summary score combines the word recall and mental status indices, ranging from 0 to 35, where higher values reflect better cognitive health. The development of these measures was coordinated by the HRS Health Working Group and informed by the expertise of cognitive psychologists, gerontologists, geriatricians, and psychiatrists. The summary scores are constructed by counting the number of correct responses across all cognitive measures included in the index (Ofstedal, Fisher, and Herzog, 2005).

In addition, we include a health variable related to fluid intelligence, assessed in the HRS using a number series test adapted from the Woodcock-Johnson (WJ-R) battery, which practitioners use to assess cognitive ability. This test evaluates fluid reasoning based on numerical patterns and mathematical relationships. Respondents are given a sequence of numbers with a missing value and must identify the pattern to determine the missing number. The test is composed of sets categorized by difficulty level. The total score reflects both the difficulty of the sets and the number of correct responses,

making it comparable to the W-scores used in the Woodcock-Johnson III (WJIII) battery tests of cognitive ability. In the WJIII norming sample, these W-scores are scaled to a mean of 500, with higher scores indicating better performance on the number series task.

To construct socioeconomic ranks, we use respondents' and their spouses' total income from the RAND data, including earnings, pensions and annuities, Social Security Income and Social Security Disability, Social Security retirement, unemployment and workers' compensation, other government transfers, household capital income, and other income. Additionally, we incorporate the income of other household members, derived from raw HRS data, including both job and non-job income. Household income is then adjusted for household size to ensure comparability across different household compositions.⁴ The rank of individuals is calculated separately across each region defined in Subsection 2.5.1.

2.4.2 Sample restriction and descriptive statistics

We restrict our sample to individuals born in the United States in 1964 or earlier. Unlike number series score-related questions asked of all respondents, cognition-related questions are asked only of individuals aged 65 or older or of new respondents. After excluding all observations with missing values in any of the variables used in the model, our final sample consists of 10,285 observations for the cognition summary score and 13,338 for the number series score.

Table 2.1 presents weighted summary statistics for the variables used in this research. In addition to sociodemographic characteristics used as controls, we include cardiovascular conditions, which have been proven to be risk factors for developing dementia. Analysing the table shows differences in the average of sociodemographic characteristics and health condition variables of individuals living in Jim Crow and non-Jim Crow regions.

In fact, Black individuals living in the non-Jim Crow region are notably younger, by around seven years on average. This age difference likely reflects historical migration

⁴Specifically, we divide total household income by the square root of the number of household members.

patterns, as Black individuals born outside the South are more often second-generation descendants of young migrants from the Great Migration, who relocated in search of better opportunities (Alexander et al., 2017; Collins, 2021; Shi et al., 2022). Consequently, the Black population in the non-Jim Crow region is disproportionately composed of younger cohorts.⁵

This younger demographic is also reflected in the smaller proportion of retired individuals among this category (47.2%). On the other hand, the number of years of education is relatively lower in the Jim Crow region, particularly for Black individuals. Regarding health conditions, the Jim Crow region exhibits worse health outcomes compared to the Non-Jim Crow region. For conditions such as high blood pressure and diabetes, disparities are even more pronounced among Black individuals compared to Non-Blacks.

The observed differences between the Jim Crow and Non-Jim Crow regions highlight the need to adjust for observable characteristics when making comparisons. To address this, in addition to controlling for these characteristics in our regressions, we implemented a reweighting approach to balance them across the two regions, as detailed in Subsection 2.5.2.

2.5 Empirical strategy

2.5.1 Econometric models

Our empirical strategy aims to isolate the effect of de facto versus de jure segregation on the absolute and the relative concentration curves of cognitive health of Black Americans. In this setting, the treatment affects the joint distribution of health and income and, consequently, their associated distributional statistics. We exploit a natural experiment within the United States: Jim Crow laws were restricted to Southern states, while states outside the South did not implement such laws. This geographical variation enables a comparison between individuals who were exposed to the Jim Crow legislation and those who were not.

⁵Since the age is a primary factor affecting cognitive abilities, a robustness check is performed in Section 2.7.

We define the Jim Crow region as the states where Jim Crow laws were in effect: Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia, and the District of Columbia (Figure 2.1). All the remaining states constitute the Non-Jim Crow group. Due to data limitations, we can only identify the place of birth at the census division level.⁶ Consequently, we are unable to isolate Missouri from the Non-Jim Crow group, which will potentially bias our estimates downward. In this context, our estimated treatment effect should be interpreted as a lower-bound estimate. Accordingly, the rank of individuals is calculated separately for each region, including all individuals born in that region.

Our identification strategy assumes that the Non-Jim Crow region provides an appropriate counterfactual for the Jim Crow region. While de jure segregation was specific to the South, de facto segregation and racial segregation were prevalent in other regions as well. During the Great Migration, many Blacks moved northward, but as James Baldwin wrote, “they do not escape Jim Crow; they merely encounter another, no less deadly variety.” The presence of racial discrimination in the North was partly a reaction of Northern whites to the large influx of African Americans into Northern cities.

In the North, segregation was less formal but deeply entrenched through private behaviour, market practices, and public policies (Sugrue, 2008). Northern Blacks lived in declining neighborhoods, attended de facto segregated schools, and faced barriers in the labour market (Hirsch, 1983; Sokol, 2014; Sugrue, 2008). Thus, while the legal context differed, both regions reflect a broader landscape of segregation. Therefore, the difference between South and North is not segregation vs. no segregation, but rather legal segregation vs. informal segregation.

To identify the causal effect of de facto segregation versus de jure segregation, we compare the differences between Black individuals and non-Black individuals within both Jim Crow and Non-Jim Crow groups. We refer to individuals as Black if they

⁶The RAND data provides nine census divisions: New England Division, Middle Atlantic Division, East North Central Division, West North Central Division, South Atlantic Division, East South Central Division, West South Central Division, Mountain Division, and Pacific Division

self-identify as Black or African American, and as non-Black if they identify as white or another racial group.⁷

As our analysis relies on a DiD framework without a time dimension, the parallel trends assumption cannot be directly tested. Instead, identification rests on the assumption that the cognitive health gap between Black and non-Black individuals born in the non-Jim Crow region would have been the same for those born in the Jim Crow region had Jim Crow laws also been implemented there. While Jim Crow and non-Jim Crow regions may differ in investments in health, education, or other determinants of cognitive outcomes, such factors are expected to affect Black and non-Black individuals similarly in the absence of racial segregation or discrimination. Finally, both race and place of birth are exogenous characteristics that individuals cannot choose, which strengthens the plausibility of our identifying assumptions.

Thus, our models include a dummy variable $black_i$ equal to 1 if individual i is Black and 0 otherwise, a dummy variable $NonJimCrow_i$ equal to 1 if individual i was born in the Non-Jim Crow region and 0 otherwise, and the interaction term $black_i \times NonJimCrow_i$ in addition to a vector of demographic characteristics X_i that includes age, gender (whether the individual is male), years of education, marital status (whether the individual is married), retirement status (whether the individual is retired), and health conditions (whether the individual has ever had diabetes, high blood pressure, or heart problems).

In fact, Jim Crow laws may have affected individual characteristics such as education and physical health, which in turn influence cognitive health. Controlling for these variables is therefore likely to underestimate the total treatment effect, as the estimated coefficient captures only the direct effect of the laws on cognitive health. For this reason, we also estimate a model without covariates.

As a baseline, we first estimate the commonly used DiD approach for each of the four health outcomes h_i as expressed in Equation 2.3.

$$h_i = \alpha + \beta black_i + \gamma NonJimCrow_i + \tau black_i \times NonJimCrow_i + X_i \psi + \epsilon_i. \quad (2.3)$$

⁷Other races in the HRS data include American Indian, Alaskan Native, Asian, Native Hawaiian, and Pacific Islander.

Since we are interested in assessing the heterogeneous treatment effect across all socioeconomic ranks q , we estimate a series of distributional DiD RIF regressions on a grid of points $q \in 0.05, 0.10, 0.15, \dots, 0.90, 0.95, 1$ as expressed in Equation 2.4:⁸

$$\begin{aligned} RIF_{C_i}^q &= \alpha(q) + \beta(q)black_i + \gamma(q)NonJimCrow_i \\ &\quad + \tau(q)black_i \times NonJimCrow_i + X_i\psi(q) + \epsilon_i, \end{aligned} \quad (2.4)$$

where $RIF_{C_i}^q$ are the values of $RIF(h_i, y_i, C_A(q, F_{HY}))$ or $RIF(h_i, y_i, C_R(q, F_{HY}))$ at observation i respectively.

The parameters $\tau(q)$ associated with the interaction terms capture the treatment effect across the distribution of socioeconomic ranks. This approach allows us to construct a treatment effect curve, providing insight into the heterogeneity of the effect across the income distribution. Note that $\tau(1)$ related to the absolute concentration curve equation at rank $q = 1$ corresponds to the average DiD treatment effect τ in Equation 2.3.

2.5.2 Reweighting approach

As discussed in Subsection 2.4.2, Jim Crow and non-Jim Crow regions exhibit imbalances in observed characteristics. To address this issue and to minimize the linear approximation errors inherent in the RIF approach, as explained in Subsection 2.3.2, we implement a reweighting strategy that adjusts the distribution of characteristics in one region to match the distribution of covariates X in the other region.⁹ This reweighting is applied in addition to controlling for these characteristics in the regressions, including race, age, gender, years of education, marital status, retirement status, and health conditions.

Using a logit model, we estimate the probability of being born in a Non-Jim Crow region as a function of demographic characteristics X , including race (Black), age, gender, years of education, marital status, labor force status, and health conditions. The predicted probabilities allow us to compute two sets of weights: the first set corresponds to the reweighted observations in the Non-Jim Crow region, adjusted

⁸Since, by construction, $C_R(1; F_{HY}) = 1$, we do not estimate the model for the relative concentration curve at rank 1.

⁹For a detailed discussion of the rationale underlying the reweighting approach, see Bchi, Makdissi, and Yazbeck (2024).

to match the distribution of characteristics in the Jim Crow region. The second set corresponds to the reweighted observations in the Jim Crow region, adjusted to match the distribution of characteristics in the Non-Jim Crow region.

To estimate the weights, we pool the data from both regions and compute the reweighting factors $\Psi_1(x_i)$ for observations in the Jim Crow region using the predicted probabilities as shown in equation 2.5:

$$\widehat{\Psi}_1(x_i) = \frac{\widehat{\Pr}[D = 1|X = x_i]/\widehat{\Pr}[D = 1]}{\widehat{\Pr}[D = 0|X = x_i]/\widehat{\Pr}[D = 0]}, \quad (2.5)$$

where D is a binary variable equal to 1 for individuals born in a Non-Jim Crow state and 0 otherwise.

We can proceed similarly to estimate the reweighting factors $\Psi_0(x_i)$ for observations in the Non-Jim Crow region to match the distribution of characteristics in the Jim Crow region.

2.6 Results

In this section, we present the results of estimating equations 2.3 and 2.4 for the treatment effect on the treated. However, the curves of the treatment effect on the untreated are presented in the appendix (Figure 2.28 and 2.29).

2.6.1 Descriptive statistics of the outcome variables

As shown in Table 2.1, the average cognitive summary score is higher for individuals born in non-Jim Crow states (23.33 for non-Blacks and 21.70 for Blacks) compared to those born in Jim Crow states (22.58 for non-Blacks and 19.29 for Blacks). Similarly, the mental status score and the word recall score exhibit lower levels among individuals born in Jim Crow states. The average number series score follows this pattern as well, with respondents born in non-Jim Crow states scoring higher (530.18 for non-Blacks and 518.48 for Blacks) compared to those born in Jim Crow states (522.29 for non-Blacks and 499.44 for Blacks)

Figures 2.5, 2.6, 2.7, and 2.8 plot the absolute and relative concentration curves for the four cognitive health outcomes, providing a visual representation of health distribution

along socioeconomic ranks. In all figures on the left, the absolute concentration curves for the Jim Crow region lie below the curves for the non-Jim Crow region at any rank, indicating that individuals born in non-Jim Crow states cognitively outperform those born in Jim Crow states across all socioeconomic ranks. At social rank 1, the absolute concentration curve corresponds to the population's average health outcome.

Regarding health inequality, all the curves on the right lie below the 45-degree line, indicating that wealthier individuals perform better, although the disparity is not substantial. Furthermore, the curves for the Jim Crow region lie below those for the non-Jim Crow region for the cognition summary score and word recall score, suggesting greater health inequality in the Jim Crow region, although the difference is very small. However, for the word recall score and number series score, the curves for the two regions overlap, indicating no significant difference in health inequality between these regions.

2.6.2 Estimated treatment effect on health achievement

Figure 2.9 and Table 2.2 illustrates selected points from the treatment effect curve on the absolute concentration curve of the cognition summary score, $\tau(q)$, estimated using the DiD RIF regression (in blue), alongside the treatment effect τ estimated from the traditional DiD regression (in red), both with their corresponding 95% confidence intervals. The graphic on the left displays the results of the treatment effect estimation without covariates. In contrast, the graphic on the right incorporates covariates, providing a clearer view of how other demographic characteristics influence the observed outcomes.

The results of the estimation without covariates and reweighted reveal a positive and significant effect across all ranks, suggesting that Black individuals born in Non-Jim Crow states consistently perform better in terms of cognitive functioning than their counterparts born in Jim Crow states. This finding highlights the significant role of structural differences between these regions in shaping long-term cognitive outcomes. When covariates such as age, gender, marital status, education, and health conditions are added to the models, the treatment effect estimated by the traditional DiD regression becomes non-significant at the 5% level. This change suggests that

these individual characteristics partially account for some of the observed differences. However, the treatment effect estimated by the DiD RIF regression remains positive and significant at the 5% level for most social ranks, except for rank $q = 1$. This result aligns with the traditional DiD regression, which also shows non-significance at the highest rank.

Figure 2.10 provides a complete view of the treatment effect curve across the entire distribution of social ranks. Our analysis focuses on the right-hand graphic, which incorporates covariates. The estimated treatment effect remains positive and significant at the 5% level for ranks up to $q = 0.85$, reflecting the consistent cognitive advantage for Black individuals born in Non-Jim Crow states compared to those born in Jim Crow states. The magnitude of the effect varies across the distribution of socioeconomic rank, ranging from 0.24 to 0.8, which corresponds to approximately 1.1% to 3.5% of the average cognition summary score (22.8). Beyond rank $q = 0.85$, the effect becomes non-significant, particularly at the upper end of the distribution. This finding suggests that the benefits of being born in a Non-Jim Crow region, while substantial, are not uniformly distributed across the population.

The non-significance of the effect at rank $q = 1$ underscores an important limitation of the traditional DiD estimation approach. By focusing on the average treatment effect across the entire population, it can obscure critical variations in effects across different social strata. The DiD RIF regression addresses this limitation by providing a more nuanced view, capturing how the treatment effect varies along the distribution of social ranks. These results reinforce the need to examine distributional effects in policy evaluation.

In terms of the mental status score, Figure 2.11 presents the treatment effect curve for the absolute concentration curve of mental status score, illustrating how the effects of being born in a Non-Jim Crow region differ across social ranks. The results show a consistently positive and significant effect across nearly all social ranks, with the effect magnitude varying slightly across the distribution, ranging from 0.08 to 0.56, corresponding to approximately 0.7% to 4.3% of the average mental status score (13.0). This indicates that Black individuals born in Non-Jim Crow regions exhibit better mental status later in life compared to their counterparts born in Jim Crow regions.

However, in terms of the word recall score, the estimation with covariates, as shown in Figure 2.12, reveals no significant treatment effect across nearly all social ranks. This suggests that the observed effects on cognitive functioning are predominantly driven by the mental status component rather than by word recall ability. The absence of significant effects in the word recall score across social distribution suggests that differences between individuals born in Non-Jim Crow and Jim Crow regions do not extend to memory-related tasks when other demographic factors are taken into account.

This finding underscores the importance of mental status in explaining the cognitive advantages for individuals born in Non-Jim Crow regions. It may indicate that factors influencing mental processing, such as education, access to healthcare, or socio-environmental conditions, play a larger role in cognitive outcomes than memory recall alone. This result highlights the need for a more nuanced understanding of how various cognitive domains are impacted by historical segregation legislation, suggesting that the broader mental functioning may have been more responsive to the environmental advantages associated with being born in Non-Jim Crow regions.

Similarly, Figure 2.13 shows a positive and significant effect for Black individuals born in the Non-Jim Crow region compared to those born in the Jim Crow region in terms of the number series score, except for the lower end of the distribution. The magnitude of the effect varies across the distribution of socioeconomic rank, ranging from -0.54 to 6.3, which corresponds to approximately 0.1% to 1.2% of the average number series score (529.5). These results further suggest that Black individuals born in the Non-Jim Crow region have an advantage in mathematical ability later in life, relative to their counterparts born in Jim Crow regions. This pattern reinforces earlier findings on cognitive functioning, indicating that the systemic advantages provided by the Non-Jim Crow environment may extend to domains of cognitive performance, such as numerical reasoning.

The positive and significant treatment effects observed across various indicators above highlight the long-term benefits for Black individuals born in regions without formalized Jim Crow laws. These advantages translate into better cognitive outcomes later in life, with the treatment effect being significant across most social ranks, suggesting

that the benefit extends beyond specific socioeconomic groups. This consistent advantage underscores the broad and lasting impact of historical racial segregation and the relative benefits of being born in the Non-Jim Crow region.

where better educational resources, reduced racial discrimination, and more favorable socio-economic conditions likely contributed to enhanced cognitive and mathematical abilities.

These findings provide important evidence on the impact of de facto versus de jure segregation, supporting the hypothesis that early-life social and environmental conditions have a lasting impact on cognitive health in later life. By focusing on the treatment effect across the entire distribution of social ranks, this analysis captures the nuanced ways in which inequalities continue to shape individual outcomes.

2.6.3 Estimated treatment effect on socioeconomic health inequality

In terms of socioeconomic health inequalities, Figures 2.14, 2.15, 2.16, and 2.17 illustrate the treatment effect on the treated for the relative concentration curves of the cognition summary score, mental status score, word recall score, and number series score, respectively. For the cognition summary score, the effect is positive and statistically significant at the 5% level. Since the relative concentration curve lies below the 45-degree line, the results indicate that being born in the Non-Jim Crow region shifts the curve upward, reducing health inequalities across nearly the entire distribution of social ranks. A similar conclusion is observed for the mental status score, reinforcing the finding that individuals born in the Non-Jim Crow region experience less disparity in cognitive functioning compared to those born in Jim Crow regions.

For the word recall and number series scores, the treatment effects also remain positive and significant at the 5% level, though with some variations. Specifically, the word recall score shows non-significant effects in the middle of the distribution, while the number series score becomes non-significant at the bottom of the distribution. Despite these localized exceptions, the results suggest that being born in the Non-Jim Crow region contributes significantly to reducing health inequalities overall. This pattern indicates that the benefits of being born in a region extend beyond average cognitive outcomes to include a more equitable distribution of cognitive abilities.

These findings underscore the broader impact of early-life regional conditions on reducing socioeconomic disparities in cognitive health outcomes. The upward shift in the relative concentration curves highlights how the absence of Jim Crow laws has helped narrow socioeconomic health inequalities. Moreover, the persistence of significant effects across most of the distribution for multiple measures of cognitive health suggests that the advantages of being born in Non-Jim Crow regions are substantial.

2.7 Robustness check

2.7.1 Age

As shown in Table 2.1, the nearly 7-year difference in average age between Black individuals born in Non-Jim Crow and Jim Crow regions may give those from the former group an advantage in cognitive test performance, as they are generally younger. To address this concern, in addition to controlling for age, we restrict our sample to individuals aged 60-80 years, thereby reducing the average age gap between the two groups to a comparable level (Table 2.3). Figure 2.18 plots the results of the estimation of the treatment effect curves on health achievement of the four indicators. In fact, our results remain robust after the age restriction.

Regarding health inequalities, Figure 2.19 presents the results of estimating the treatment effect curves on the relative concentration curves for the four indicators. The results remain robust for the total cognition summary score and the mental status score. The treatment effects on the relative concentration curves for the word recall and number series scores remain positive but less precise. The smaller sample size likely drives this in the restricted age range.

2.7.2 Migration

The Great Migration was one of the largest movements of African Americans from Southern states to Northern, Midwestern, and Western states between the 1910s and the 1970s. Even though the place of birth—used as the treatment indicator—is an independent decision, we restricted our sample to individuals who were both born and

residing in the same region during all survey waves to avoid any potential contamination of the effect due to migration. Applying this restriction retains 77% of the sample.¹⁰

The results of the estimation presented in Figure 2.20 show that the results remain largely comparable, but less precise, particularly for the total cognition summary score.¹¹ In terms of health inequalities, Figure 2.21 shows patterns broadly consistent with the original estimates.

2.7.3 Placebo test

As an additional robustness check, we perform a placebo test by randomly assigning a treatment status to individuals in the control group (those born in the Jim Crow region). The results, displayed in Figure 2.22 for health achievement and Figure 2.23 for health inequalities, show that, apart from a few points in the distribution, none of the estimated effects on the four health outcomes are significant even at the 10% level.

Similarly, when the placebo test is applied to the treatment group (individuals born in the non-Jim Crow region), none of the effects are significant at the 10% level (Figures 2.26 and 2.27).

2.8 Conclusion

This study contributes to the growing body of literature on the long-term health consequences of segregation by providing new causal evidence on the impact of legalized segregation through Jim Crow laws on the cognitive health of older adults in the United States. Leveraging a natural experiment comparing states where Jim Crow laws were not enacted to those where they were enforced, we examined how de jure segregation shaped cognitive functioning in later life. Our focus was on Black individuals

¹⁰The HRS also asks about residence during school age (around 10 years old); 92% of respondents report living in the same region where they were born.

¹¹For the total cognition summary score, the effect remains positive, with a magnitude similar to the original estimates. While it is not statistically significant at the 5% level except at the 35th, 50th, 55th, 65th, and 70th social ranks, it is significant at least at the 10% level from the 35th to the 80th social ranks. For the number series score, the effect is positive starting from the 25th social ranks and comparable in magnitude to the original results. It is statistically significant at the 5% level from the 55th to the 80th social ranks, and at the 10% level starting 85th social rank.

born before 1964—those directly exposed to the institutionalized racial segregation of the Jim Crow era—and their cognitive health outcomes in older adulthood.

Besides the causal effect of Jim Crow laws on cognitive health, a key contribution of this research lies in its methodological innovation. While previous studies have largely relied on difference-in-differences (DiD) approaches to estimate the average treatment effects of a policy, our analysis goes beyond average effects. We use a distributional DiD RIF framework based on the recentered influence function (RIF) regression method introduced by Firpo, Fortin, and Lemieux (2009) and further extended by Bchi, Makdissi, and Yazbeck (2024) to the absolute and relative concentration curves. This framework allows us to estimate not only the average treatment effect of de facto versus de jure segregation but also to assess how these effects vary across the entire distribution of socioeconomic ranks. By examining treatment effects on both relative and absolute concentration curves for cognitive health, we can capture nuanced distributional impacts that traditional DiD models may obscure.

Empirically, we used data from the 2016 wave of the Health and Retirement Study (HRS), a nationally representative longitudinal survey of older adults in the U.S. Cognitive health outcomes were measured using the total cognition summary score, mental status score, word recall score, and number series score. These indicators have been widely validated in the literature, including through their use in the Aging, Demographics, and Memory Study (ADAMS) and their linkage to clinical diagnoses of dementia and mild cognitive impairment via the Langa–Weir classification (Crimmins et al., 2011; Langa et al., 2009).

Our findings reveal several important insights. First, at the average level, the absence of Jim Crow laws (de facto segregation) does not show a significant treatment effect on the total cognition summary score compared to de jure segregation. However, when examining the distributional effects through our distributional DiD RIF approach based on concentration curves, we find significant positive treatment effects up to the 0.85 social rank. This suggests that Black individuals from lower and middle socioeconomic backgrounds benefited the most from the absence of Jim Crow laws in terms of cognitive functioning. Similar patterns are observed for the mental status and number series scores, with positive, significant treatment effects across much of

the income distribution, though with varying magnitudes. In contrast, we observe no significant effect on the word recall score, suggesting potential heterogeneity in how different cognitive domains respond to early-life exposure to institutionalized segregation.

Second, our analysis of relative concentration curves indicates that the absence of Jim Crow laws led to a reduction in cognitive health inequalities among older adults. The upward shift in the relative concentration curves implies that cognitive health became more equitably distributed across income ranks in states that had earlier escaped Jim Crow laws. This finding underscores the role of legalized segregation as a fundamental determinant of health inequalities. It demonstrates that policy efforts to dismantle segregationist laws can yield long-term benefits for health equity.

From a policy perspective, our findings underscore the critical importance of addressing the root causes of health disparities, including the enduring effects of segregation. Dismantling discriminatory laws and institutions represents an essential first step, but additional interventions are needed to tackle the persistent socioeconomic inequalities that continue to shape health outcomes for marginalized populations. Investments in education, healthcare access, and community resources are crucial for promoting cognitive health equity and reducing the burden of dementia and cognitive impairment among aging populations.

In conclusion, this study demonstrates that the absence of Jim Crow laws had a significant and beneficial impact on the cognitive health of Black older adults in the U.S., particularly for those from disadvantaged socioeconomic backgrounds. By applying distributional DiD RIF regression, we provide a more comprehensive understanding of how policy changes affect not only average health outcomes but also health inequalities. While this study assessed the effect of legislation rather than the direct effect of segregation, it highlights that laws play an important role as public policy instruments—often easier to implement than changing individual behaviors. The effect of segregation itself, however, may be even greater.

Future research should explore the mechanisms by which early-life exposure to segregation translates into later-life cognitive health outcomes, including factors such

as educational quality, neighborhood conditions, discrimination throughout the life course, and differential access to healthcare and social resources.

Despite these contributions, several limitations warrant consideration. Although our natural experiment provides credible estimates of causal effects, future research should investigate settings in which individuals are differentially exposed to Jim Crow laws while living in the same environment. For instance, comparing individuals living in Jim Crow regions before and after 1964—when those born after 1964 are old enough to be compared with those born before—could yield additional insights into the treatment effects. Furthermore, to better understand the broader impact of segregation, similar analyses could be conducted in other countries, such as Canada or South Africa.

Appendix

TABLE 2.1: Weighted descriptive statistics

	Non-Jim Crow region		Jim Crow region	
	Non-Black	Black	Non-Black	Black
Sociodemographic characteristics				
Age	69.39 (10.83)	62.51 (9.42)	69.35 (10.50)	69.01 (10.42)
Years of education	13.78 (2.45)	13.55 (2.31)	13.12 (2.97)	12.22 (2.91)
Male	48.52	42.15	44.29	41.54
Married or partnered	63.72	40.56	63.22	41.65
Retired	67.39	47.19	68.71	69.62
Ever had high blood pressure	56.95	65.74	62.30	79.52
Ever had diabetes	21.53	28.75	24.99	36.02
Ever had heart problems	25.96	18.10	29.06	27.17
Health outcomes				
Total cognition summary score	23.33 (4.56)	21.70 (4.26)	22.58 (4.98)	19.29 (5.52)
Total mental status summary score	13.26 (2.04)	12.35 (2.20)	12.85 (2.38)	11.04 (2.98)
Word recall score	10.07 (3.34)	9.36 (2.95)	9.73 (3.41)	8.25 (3.37)
Number Series score	530.18 (27.46)	518.48 (26.65)	522.29 (31.18)	499.44 (34.87)
Household income				
	64,504.45 (84,156.64)	51,252.46 (58,322.32)	61,788.34 (101,833.23)	34,084.33 (50,864.63)
Number of observations	5,755	545	2,469	1,516

Note: Standard deviation between parentheses

Source: Authors' own estimations, HRS 2016.

TABLE 2.2: Model treatment effect on treated of $C_A(q)$ of cognition summary score

	$RIF(C_A(0.25))$	$RIF(C_A(0.50))$	$RIF(C_A(0.75))$	$RIF(C_A(1))$	DiD
Panel A					
Non-Jim Crow	0.10649** (0.05113)	0.19455*** (0.07153)	0.24633*** (0.08193)	0.20605** (0.08632)	0.20605** (0.08632)
Black	-1.44627*** (0.14526)	-2.09321*** (0.19358)	-2.87138*** (0.21580)	-3.20483*** (0.20635)	-3.20483*** (0.20635)
Non-Jim Crow × Black	0.95245*** (0.22646)	1.32694*** (0.27591)	1.75207*** (0.33248)	1.57720*** (0.33864)	1.57720*** (0.33864)
Constant	5.20290*** (0.04085)	10.84867*** (0.05605)	16.82494*** (0.07317)	23.12236*** (0.08111)	23.12236*** (0.08111)
Covariates	No	No	No	No	No
Panel B					
Non-Jim Crow	0.13862*** (0.05034)	0.25845*** (0.07064)	0.35924*** (0.08030)	0.34673*** (0.08492)	0.34673*** (0.08492)
Black	-1.30001*** (0.13484)	-1.83348*** (0.16789)	-2.38517*** (0.17358)	-2.56623*** (0.16152)	-2.56623*** (0.16152)
Non-Jim Crow × Black	0.59546*** (0.20710)	0.70295*** (0.24446)	0.78935*** (0.29458)	0.43475 (0.29713)	0.43475 (0.29713)
Constant	-0.18113*** (0.06293)	-0.51657*** (0.07688)	-0.67363*** (0.08160)	-0.77524*** (0.08616)	-0.77524*** (0.08616)
Covariates	Yes	Yes	Yes	Yes	Yes

Note: ** $p < 0.05$; *** $p < 0.01$. At $q = 1$.

The RIF regression is equivalent to the standard DiD estimator.

Covariates include age, gender, years of education, marital status, labour force status, diabetes, high blood pressure, and heart problems.

Source: Authors' own estimations, HRS 2016.

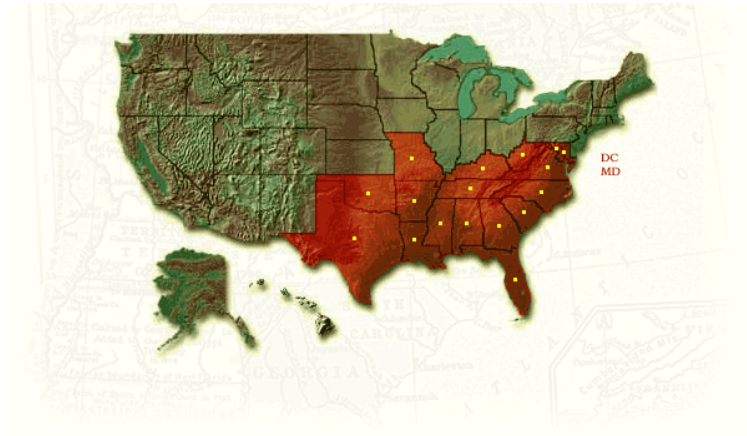
TABLE 2.3: Weighted descriptive statistics for individuals aged 60 to 80 years

	Non-Jim Crow region		Jim Crow region	
	Non-Black	Black	Non-Black	Black
Sociodemographic characteristics				
Age	71.01 (4.43)	70.11 (4.45)	71.29 (4.38)	70.86 (4.44)
Years of education	13.90 (2.32)	13.30 (2.25)	13.19 (2.84)	12.20 (2.79)
Male	47.53	40.54	42.46	39.03
Married or partnered	68.60	41.74	66.97	42.46
Retired	82.26	83.11	84.15	82.33
Ever had high blood pressure	61.07	78.54	65.79	83.44
Ever had diabetes	24.31	41.15	27.64	41.57
Ever had heart problems	27.25	24.96	31.40	28.12
Health outcomes				
Total cognition summary score	23.78 (4.35)	21.33 (4.46)	22.92 (4.80)	19.44 (5.23)
Total mental status summary score	13.40 (1.96)	12.27 (2.35)	12.95 (2.24)	11.09 (2.89)
Word recall score	10.38 (3.25)	9.05 (2.95)	9.96 (3.37)	8.35 (3.19)
Number Series score	530.33 (26.81)	514.82 (25.07)	520.99 (30.13)	494.66 (35.32)
Household income				
	67,402.97 (93,495.91)	44,707.24 (48,814.68)	60,680.21 (10,4487.16)	31,378.39 (29,884.39)
Number of observations	3,361	247	1,476	879

Note: Standard deviation between parentheses

Source: Authors' own estimations, HRS 2016.

FIGURE 2.1: Map of Jim Crow America



Source: Arthur and Emalie Gutterman Family Center for Holocaust and Human Rights Education Dorothy F. Schmidt College of Arts and Letters Florida Atlantic University

FIGURE 2.2: Relative concentration curve and concentration index

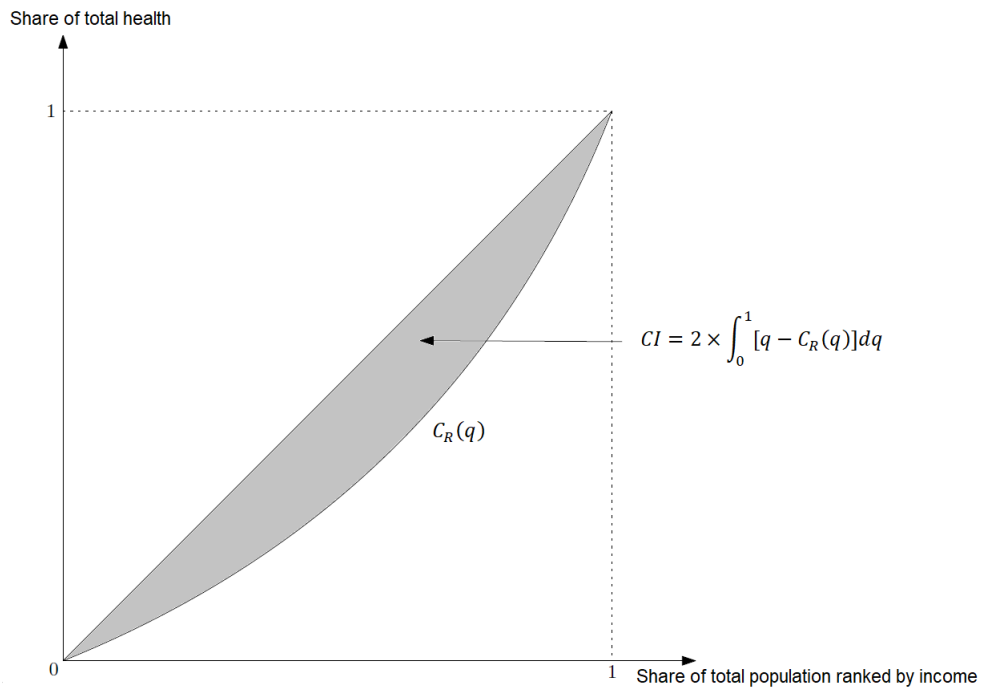


FIGURE 2.3: Absolute concentration curve

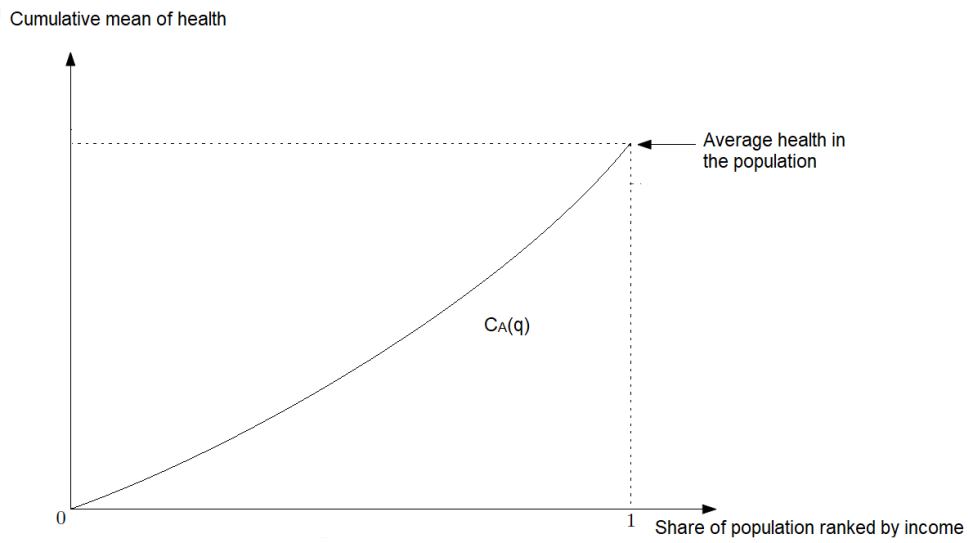


FIGURE 2.4: Total cognition summary score and its components

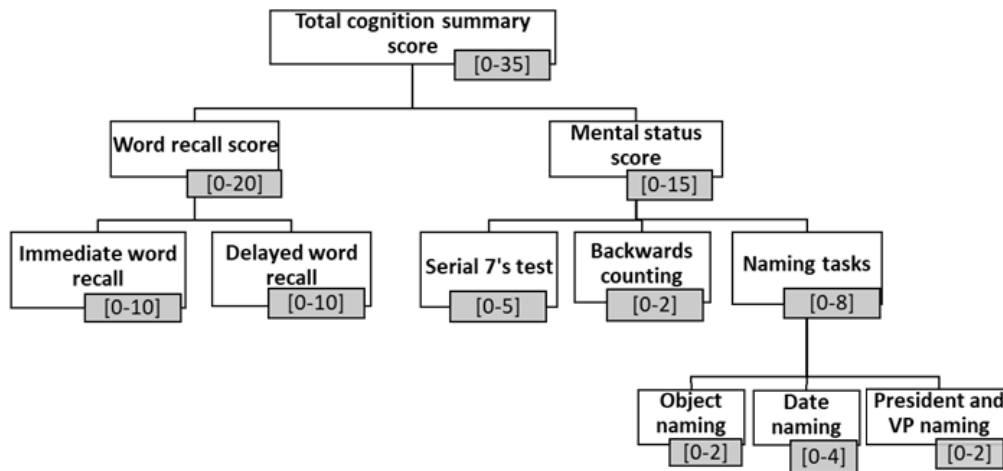
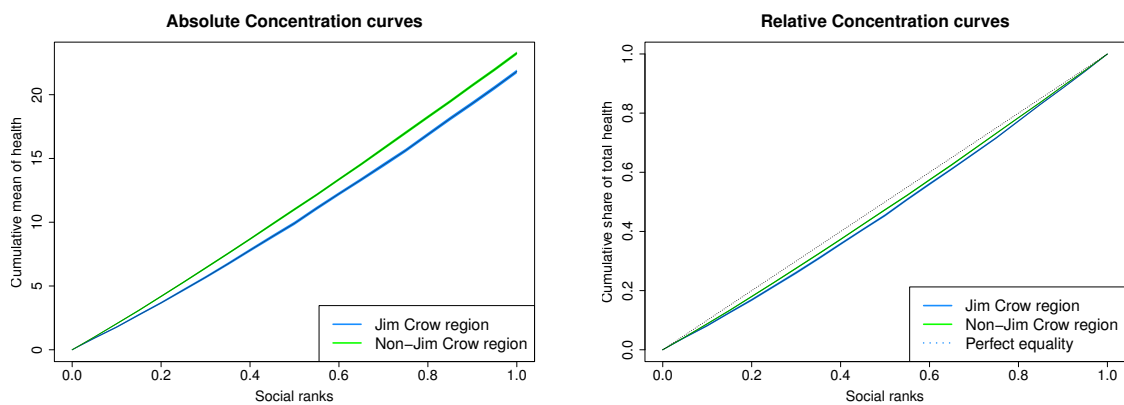
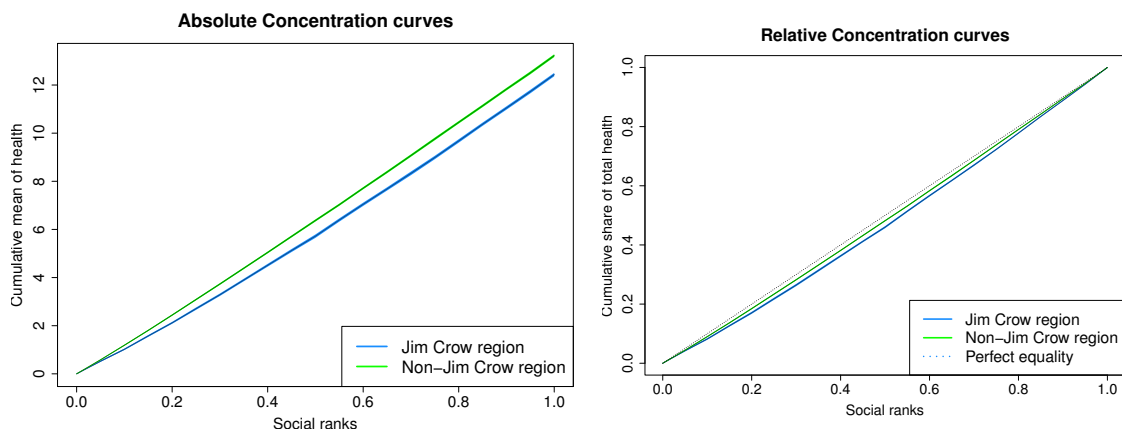


FIGURE 2.5: Absolute and relative concentration curve-Cognition score



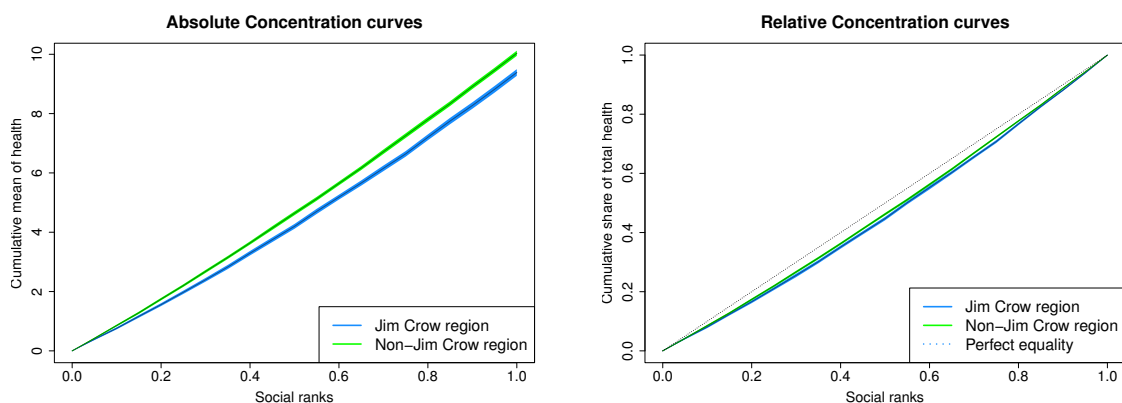
Note : The figure displays the absolute concentration curves in the left panel and the relative concentration curves in the right panel for Jim Crow and non-Jim Crow regions. In the right panel, the curve for the Jim Crow region lies below that of the non-Jim Crow region, indicating that individuals born in the non-Jim Crow region achieve higher cognitive scores than those born in the Jim Crow region. In the left panel, the Jim Crow region’s curve also lies below that of the non-Jim Crow region, suggesting that health inequalities are more pronounced in the Jim Crow region.

FIGURE 2.6: Absolute and relative concentration curve-Mental status score



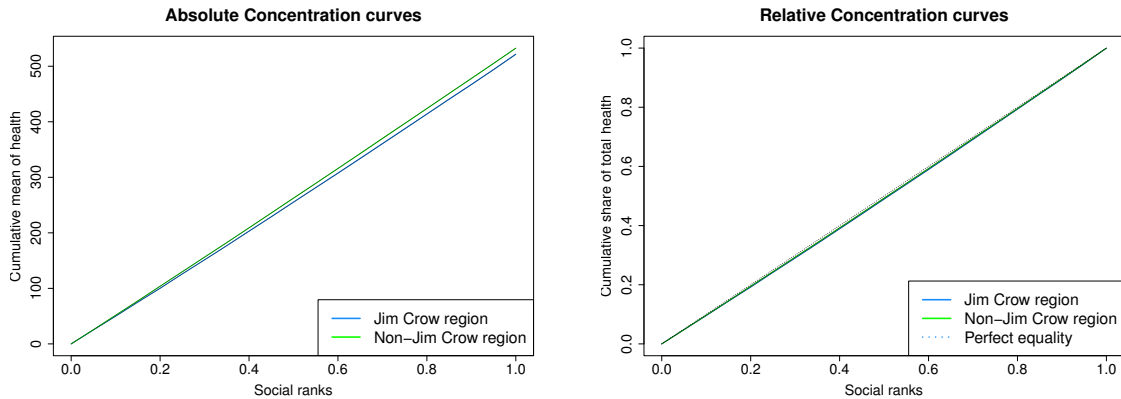
Note : The figure displays the absolute concentration curves in the left panel and the relative concentration curves in the right panel for Jim Crow and the non-Jim Crow region. In the right panel, the curve for the Jim Crow region lies below that of the Non-Jim Crow region, indicating that individuals born in the Non-Jim Crow region achieve higher mental status scores than those born in the Jim Crow region. In the left panel, the Jim Crow region's curve also lies below that of the Non-Jim Crow region, suggesting that health inequalities are more pronounced in the Jim Crow region.

FIGURE 2.7: Absolute and relative concentration curve-Word recall score



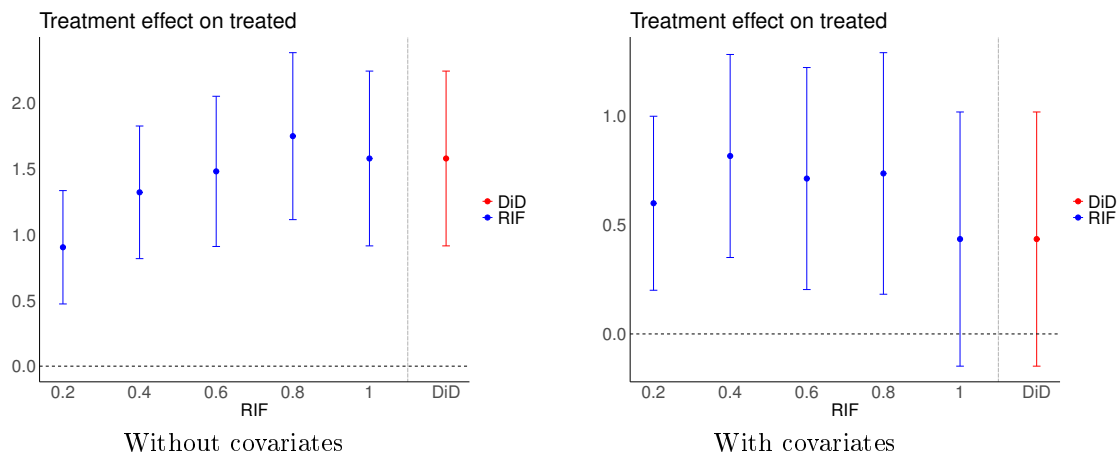
Note : The figure displays the absolute concentration curves in the left panel and the relative concentration curves in the right panel for Jim Crow and non-Jim Crow regions. In the right panel, the curve for the Jim Crow region lies below that of the Non-Jim Crow region, indicating that individuals born in the Non-Jim Crow region achieve higher word recall scores than those born in the Jim Crow region. In the left panel, the Jim Crow region's curve also lies below that of the Non-Jim Crow region, suggesting that health inequalities are more pronounced in the Jim Crow region.

FIGURE 2.8: Absolute and relative concentration curve-Number series score



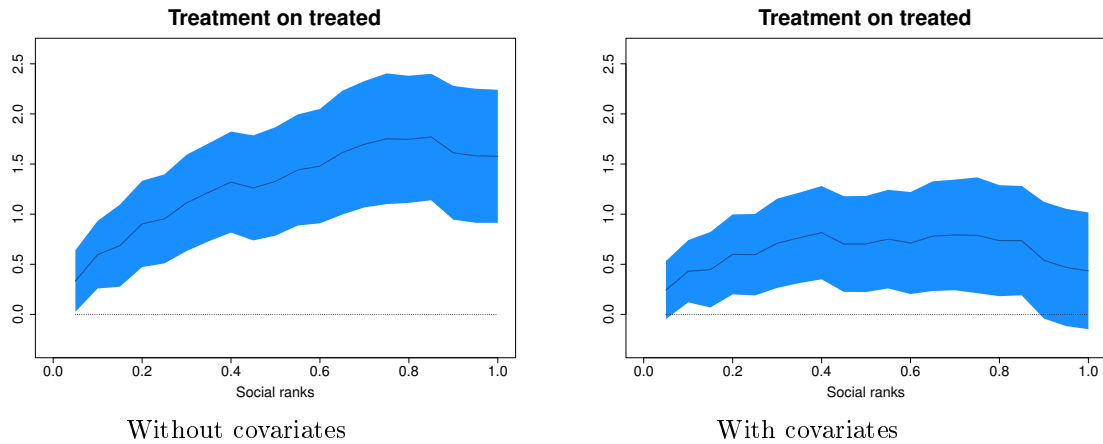
Note : The figure presents the absolute concentration curves in the left panel and the relative concentration curves in the right panel for Jim Crow and Non-Jim Crow regions. In the right panel, the curve for the Jim Crow region lies below that of the Non-Jim Crow region, indicating that individuals born in the Non-Jim Crow region achieve higher number series scores than those born in the Jim Crow region. In the left panel, the Jim Crow region’s curve also lies below that of the Non-Jim Crow region, suggesting that health inequalities are more pronounced in the Jim Crow region.

FIGURE 2.9: Treatment effect on the treated on the absolute concentration curve of Cognition summary score



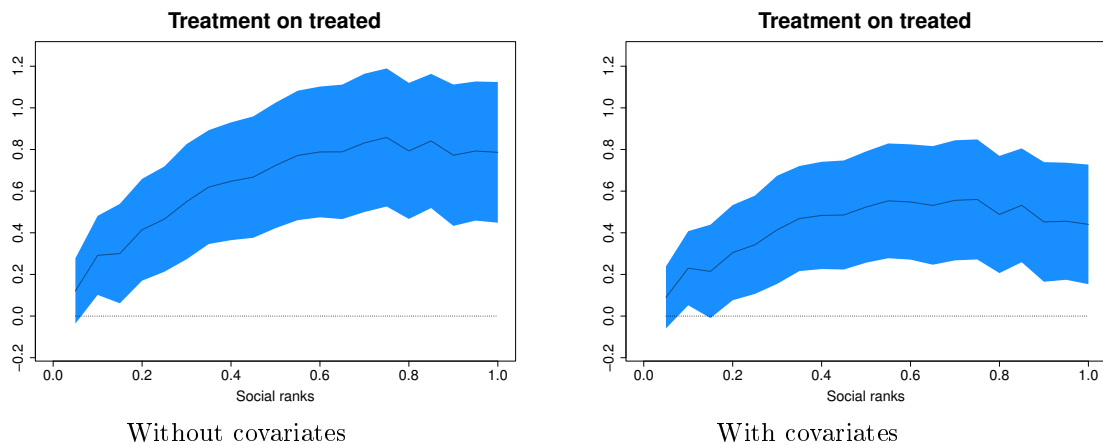
Note : The figure displays selected points of the treatment effect curve on the absolute concentration curve of the cognition summary score, together with their 95% confidence intervals, estimated using the distributional DiD RIF regression (in blue), alongside the average DiD treatment effect (in red). The right panel presents results from regressions that include covariates, whereas the left panel reports estimates from regressions without covariates. The estimate at rank 1 (blue) corresponds to the standard DiD regression. In the right panel, the average treatment effect is not statistically significant; however, for socioeconomic ranks below the 85th percentile, the treatment effect is positive and statistically significant.

FIGURE 2.10: Treatment effect on the treated for the absolute concentration curve of Cognition summary score



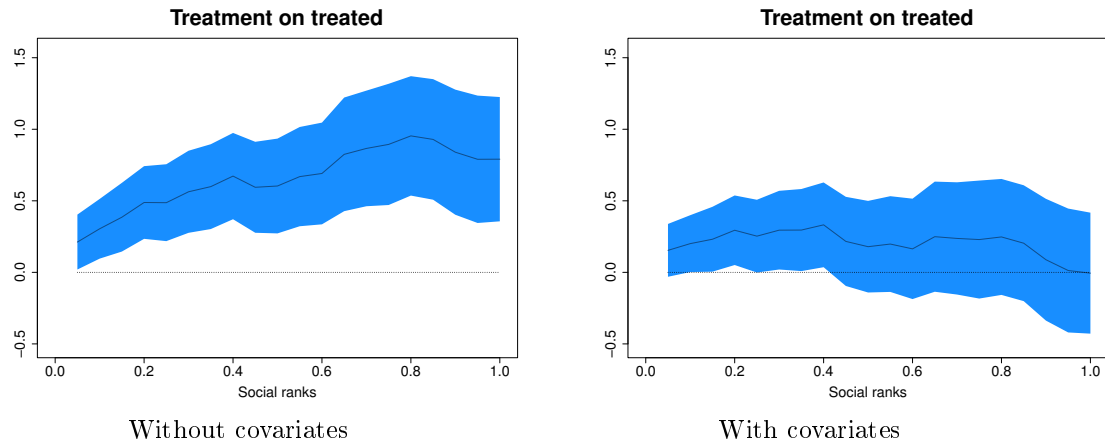
Note : The figure presents the entire treatment effect curve on the absolute concentration curve of the cognition summary score, along with its 95% confidence intervals, estimated using the distributional DiD RIF regression. The right panel reports results from regressions that include covariates, whereas the left panel shows estimates from regressions without covariates. The estimate at rank 1 corresponds to the average treatment effect. In the right panel, the average treatment effect is not statistically significant; however, for socioeconomic ranks below the 85th percentile, the treatment effect is positive and statistically significant.

FIGURE 2.11: Treatment effect on the treated for the absolute concentration curve of mental status score



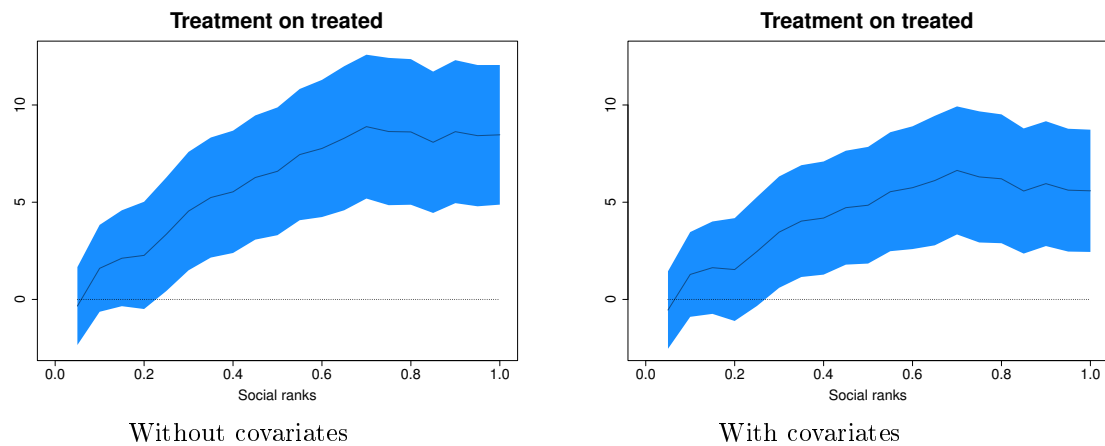
Note : The figure presents the treatment effect curve on the absolute concentration curve of the mental status score, together with its 95% confidence intervals, estimated using the distributional DiD RIF regression. The right panel reports results from regressions that include covariates, whereas the left panel shows estimates from regressions without covariates. The estimate at rank 1 corresponds to the average treatment effect. In both panels, the treatment effect is positive and statistically significant across the distribution, including the average treatment effect.

FIGURE 2.12: Treatment effect on the treated for the absolute concentration curve of word recall score



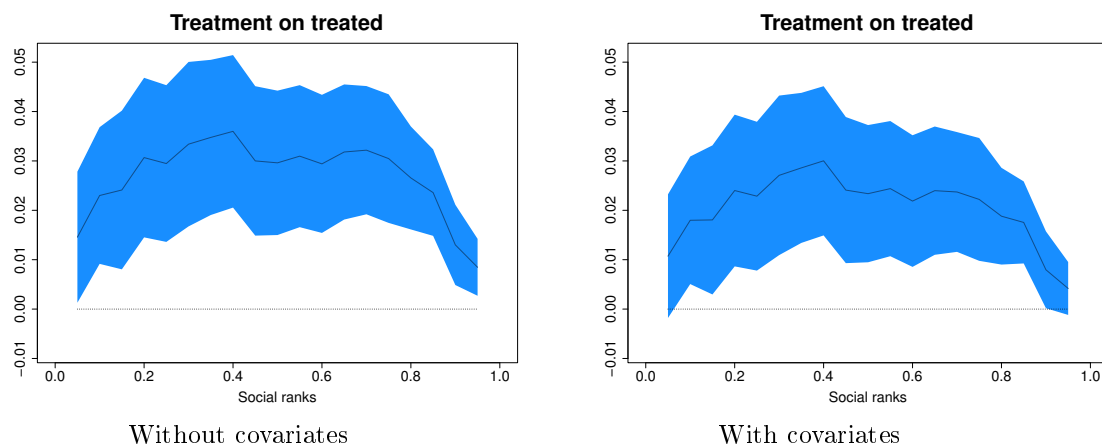
Note : The figure presents the treatment effect curve on the absolute concentration curve of the word recall score, along with its 95% confidence intervals, estimated using the distributional DiD RIF regression. The right panel reports estimates from regressions that include covariates, whereas the left panel shows results from regressions without covariates. The estimate at rank 1 corresponds to the average treatment effect. In the left panel, the treatment effect is positive and statistically significant across the distribution, including the average treatment effect. In contrast, once covariates are included (right panel), the treatment effect becomes statistically insignificant over most of the distribution.

FIGURE 2.13: Treatment effect on the treated for the absolute concentration curve of number series score



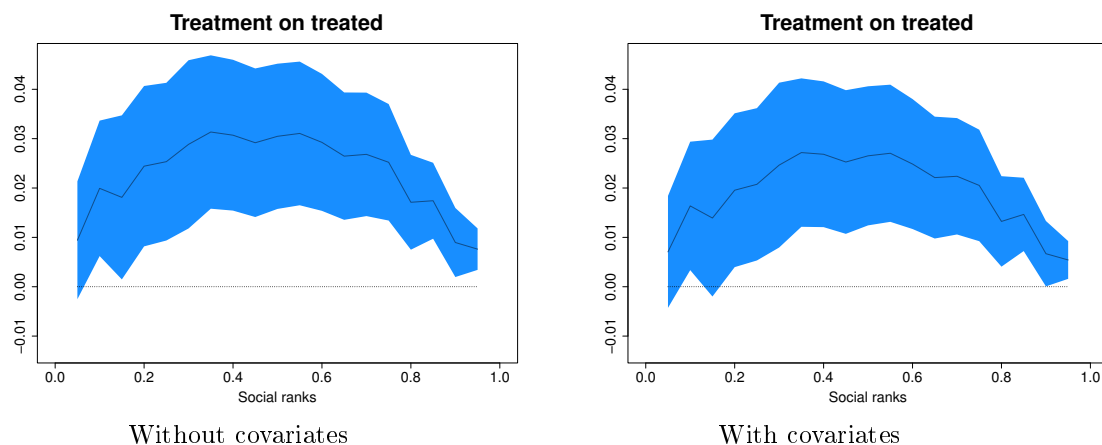
Note : The figure presents the treatment effect curve on the absolute concentration curve of the number series score, together with its 95% confidence intervals, estimated using the distributional DiD RIF regression. The right panel reports results from regressions that include covariates, whereas the left panel shows estimates from regressions without covariates. The estimate at rank 1 corresponds to the average treatment effect. In both panels, the treatment effect is positive and statistically significant across the distribution, including the average treatment effect.

FIGURE 2.14: Treatment effect on the treated for the relative concentration curve of cognition score



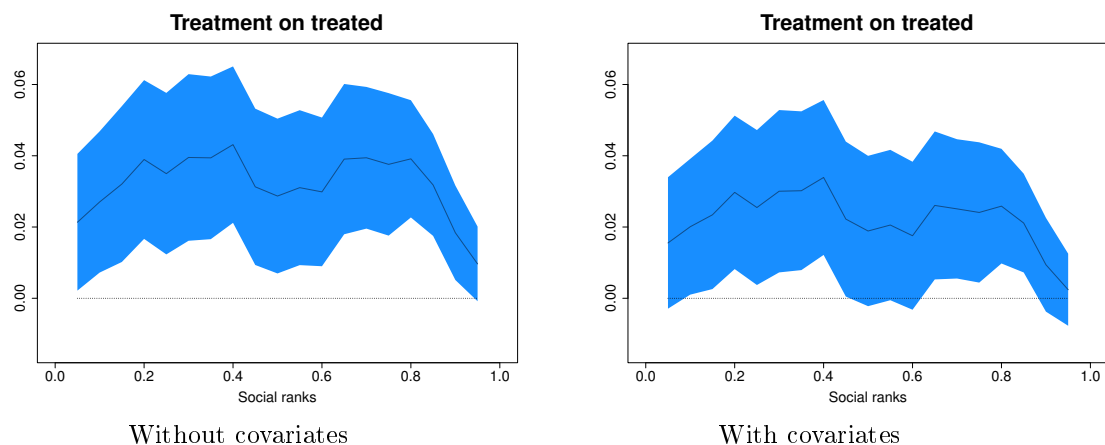
Note : The figure presents the treatment effect curve on the relative concentration curve of the cognition summary score, along with its 95% confidence intervals, estimated using the distributional DiD RIF regression. The right panel reports estimates from regressions that include covariates, whereas the left panel shows results from regressions without covariates. In both panels, the treatment effect is positive and statistically significant at nearly all points of the distribution, indicating an upward shift of the concentration curves and, consequently, a reduction in health inequalities.

FIGURE 2.15: Treatment effect on the treated for the relative concentration curve of mental status score



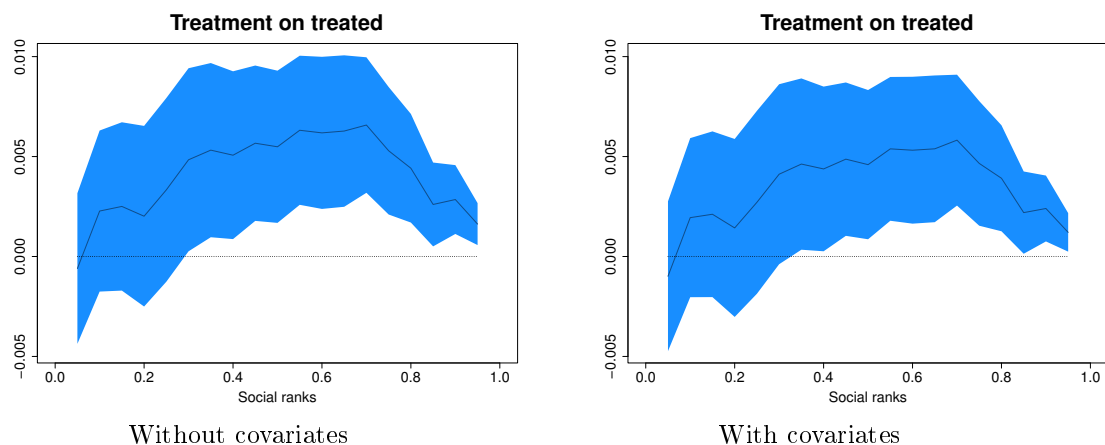
Note : The figure presents the treatment effect curve on the relative concentration curve of the mental status score, together with its 95% confidence intervals, estimated using the distributional DiD RIF regression. The right panel reports results from regressions that include covariates, whereas the left panel shows estimates from regressions without covariates. In both panels, the treatment effect is positive and statistically significant at nearly all points in the distribution, indicating an upward shift of the concentration curves and a corresponding reduction in health inequalities.

FIGURE 2.16: Treatment effect on the treated for the relative concentration curve of word recall score



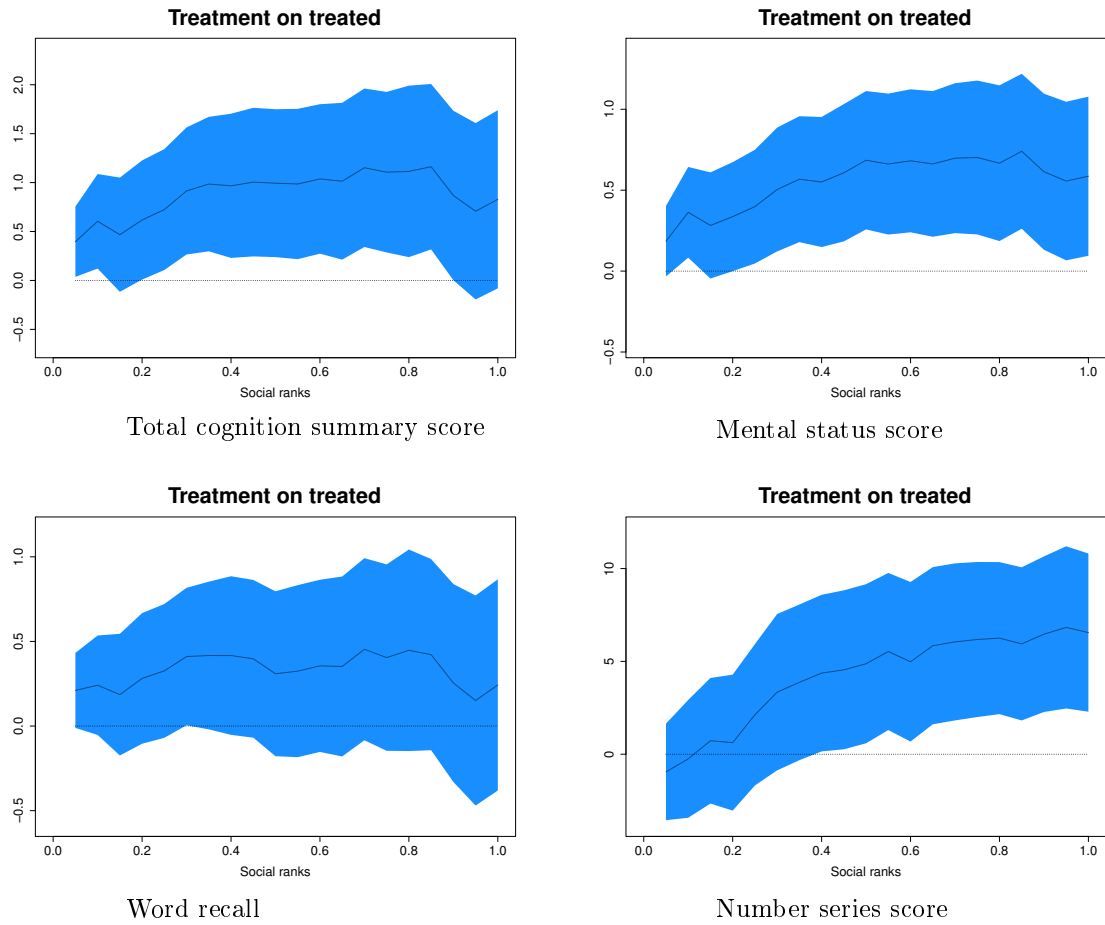
Note : The figure presents the treatment effect curve on the relative concentration curve of the word recall score, along with its 95% confidence intervals, estimated using the distributional DiD RIF regression. The right panel reports results from regressions that include covariates, whereas the left panel shows estimates from regressions without covariates. In both panels, the treatment effect is positive and statistically significant at nearly all points of the distribution, indicating an upward shift of the concentration curves and a corresponding reduction in health inequalities.

FIGURE 2.17: Treatment effect on the treated for the relative concentration curve of number series score



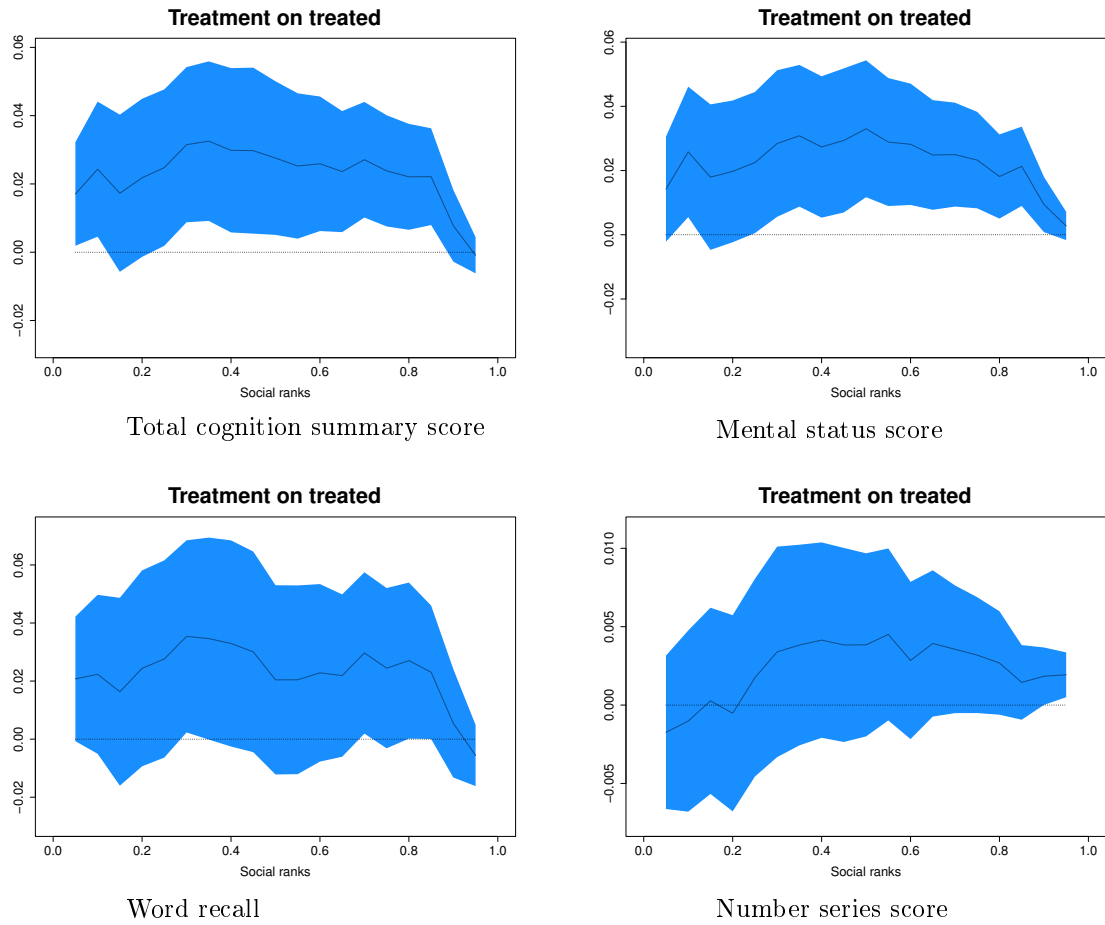
Note : The figure presents the treatment effect curve on the relative concentration curve of the number series score, together with its 95% confidence intervals, estimated using the distributional DiD RIF regression. The right panel reports results from regressions that include covariates, whereas the left panel shows estimates from regressions without covariates. In both panels, the treatment effect is positive and statistically significant at nearly all points of the distribution, indicating an upward shift of the concentration curves and a corresponding reduction in health inequalities.

FIGURE 2.18: Treatment effect on the treated for the absolute concentration curve-Age 60-80 years



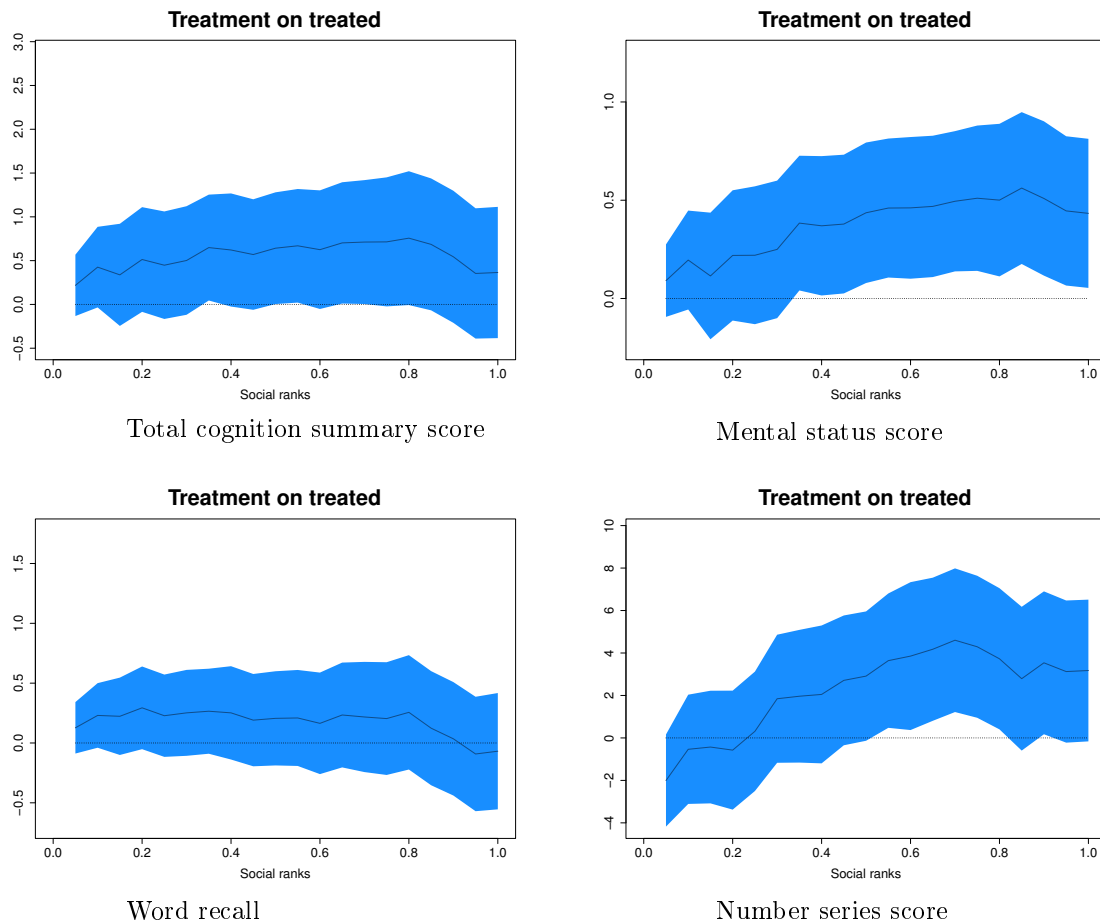
Note : The figure presents the treatment effect curves on the absolute concentration curves of the four indicators, along with their 95% confidence intervals, estimated using the distributional DiD RIF regression. The sample is restricted to individuals aged 60 to 80 years, and the regression equations include covariates. For all four indicators, the results are consistent with the original estimations.

FIGURE 2.19: Treatment effect on the treated for the relative concentration curve-Age 60-80 years



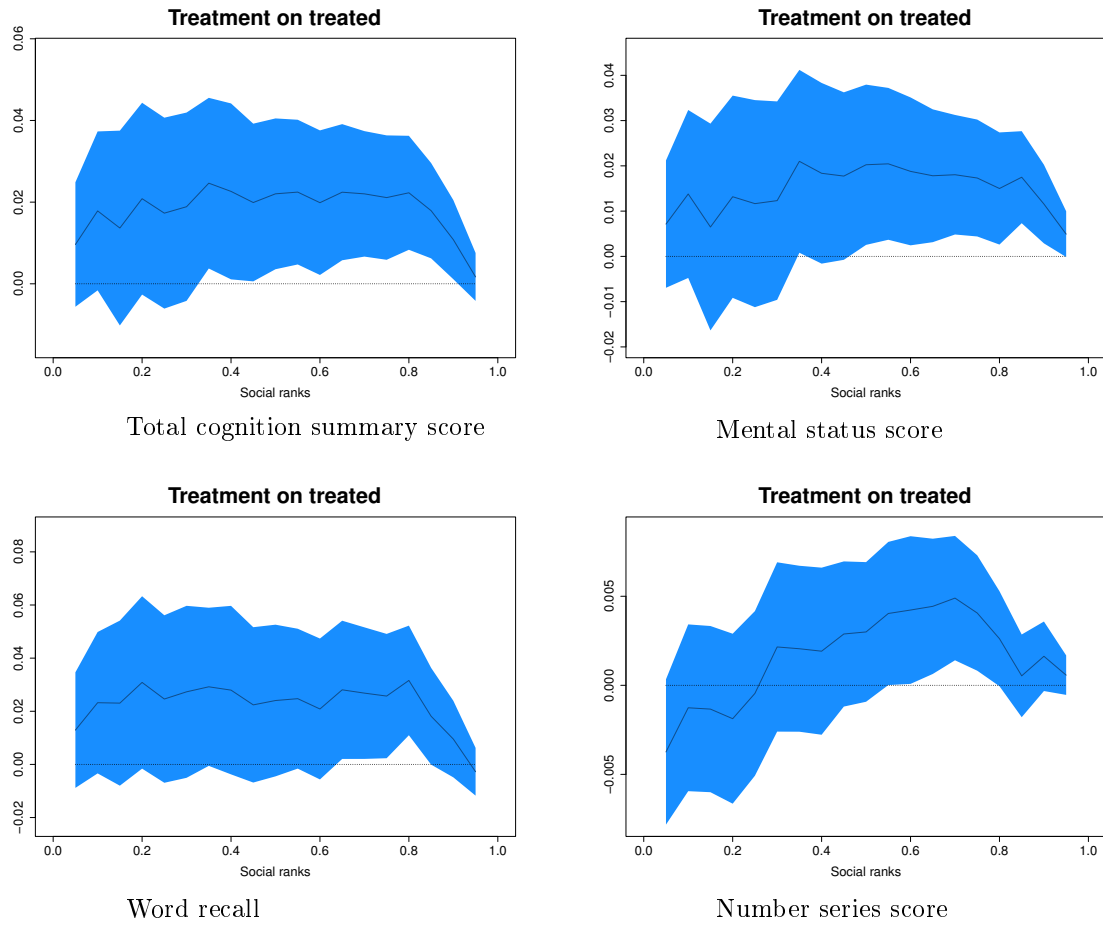
Note : The figure plots the curve of the treatment effect curve on the relative concentration curves of the four indicators and their 95% confidence intervals from the distributional DiD RIF regression. The sample is restricted to individuals aged 60 to 80 years. The regression equations include covariates. For the four indicators, the conclusions remain the same as the original estimations.

FIGURE 2.20: Treatment effect on the treated for the absolute concentration curve-No migration (the same region JC all waves)



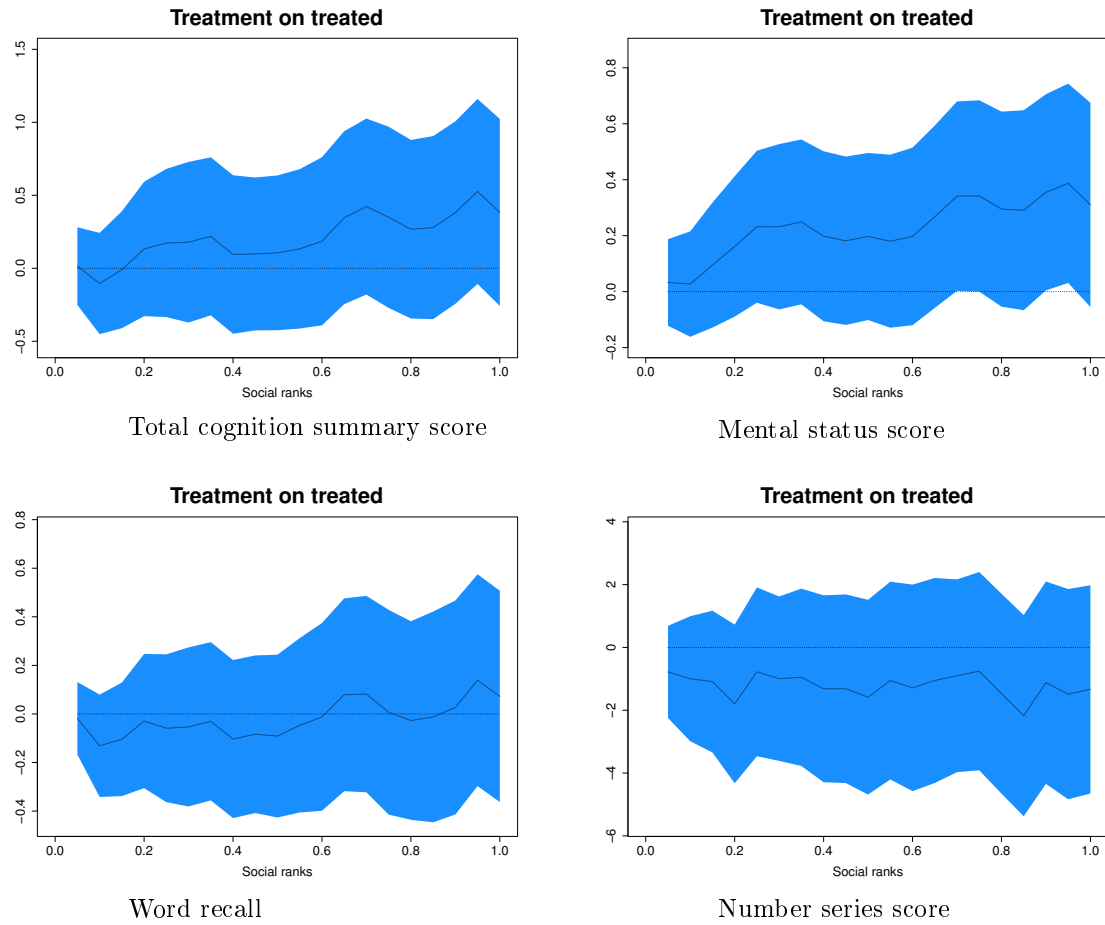
Note : The figure plots the curve of the treatment effect curve on the absolute concentration curves of the four indicators and their 95% confidence intervals from the distributional DiD RIF regression. The sample is restricted to individuals who reside in the same region of birth during all waves of HRS. The regression equations include covariates. For the four indicators, the conclusions remain the same as the original estimations, but with less power.

FIGURE 2.21: Treatment effect on the treated for the relative concentration curve-No migration (the same region JC all waves)



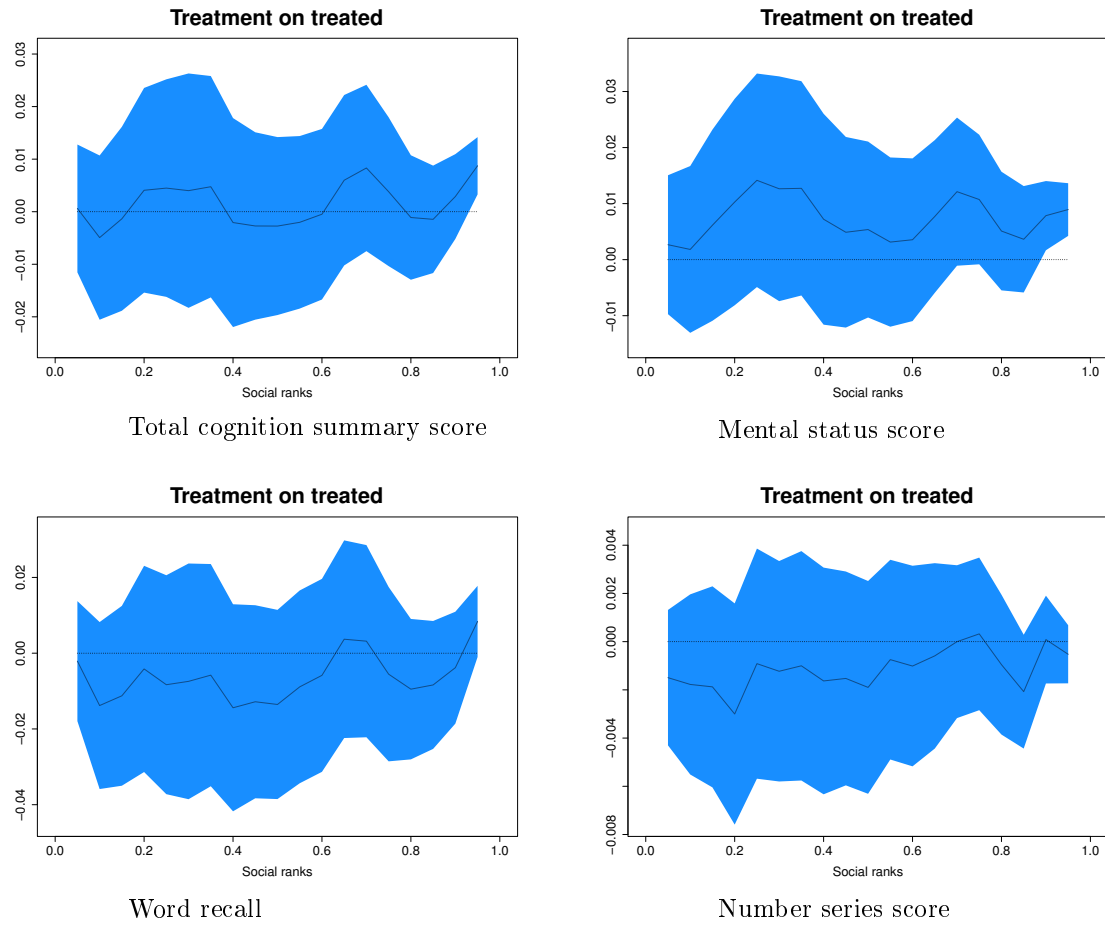
Note: The figure plots the curve of the treatment effect curve on the relative concentration curves of the four indicators and their 95% confidence intervals from the distributional DiD RIF regression. The sample is restricted to individuals who reside in the same region of birth during all waves of HRS. The regression equations include covariates. For the four indicators, the conclusions remain the same as the original estimations, but with less power.

FIGURE 2.22: Treatment effect on the treated for the absolute concentration curve-Placebo in the control group



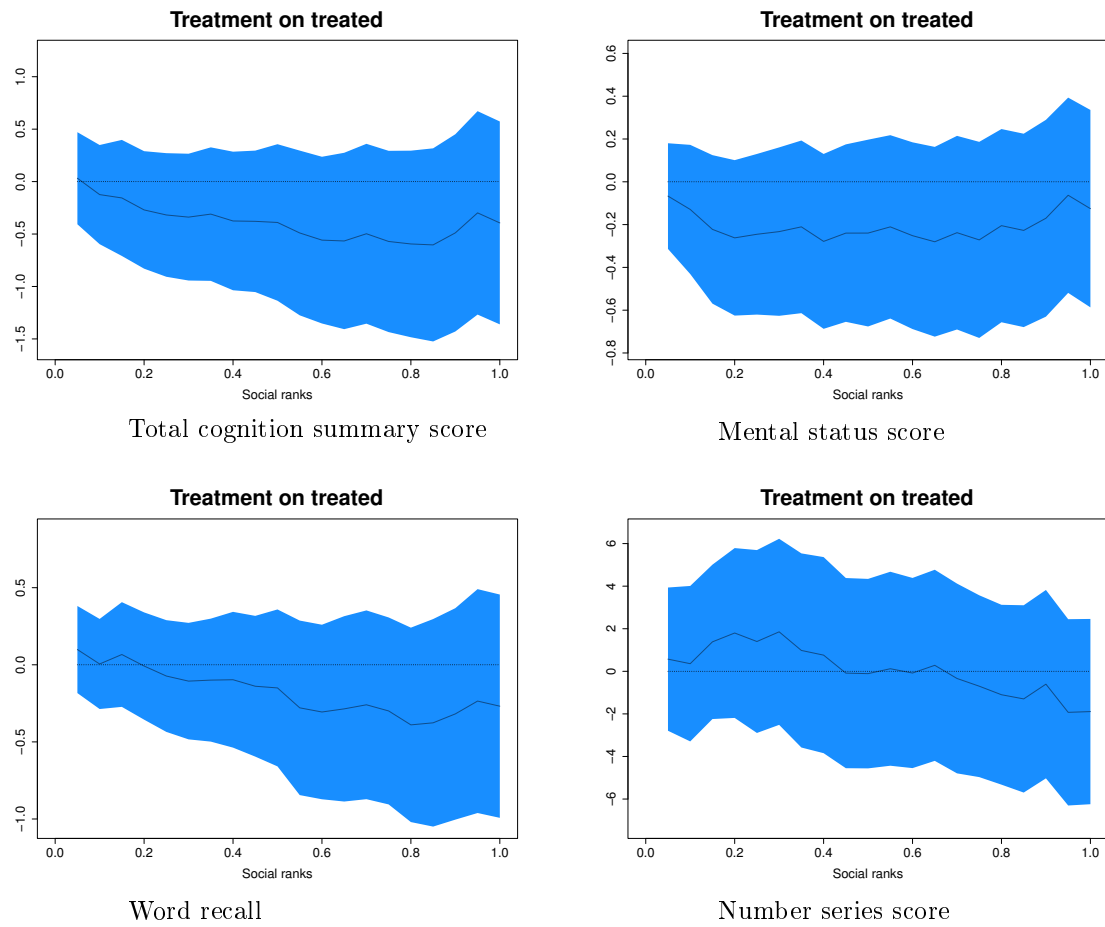
Note : The figure presents the treatment effect curves on the absolute concentration curves of the four indicators, along with their 95% confidence intervals, estimated using the distributional DiD RIF regression. The sample is restricted to individuals in the control group (Jim Crow region), and the treatment is randomly assigned. The regression equations include covariates. For all four indicators, the treatment effect is not statistically significant.

FIGURE 2.23: Treatment effect on the treated for the relative concentration curve-Placebo in the control group



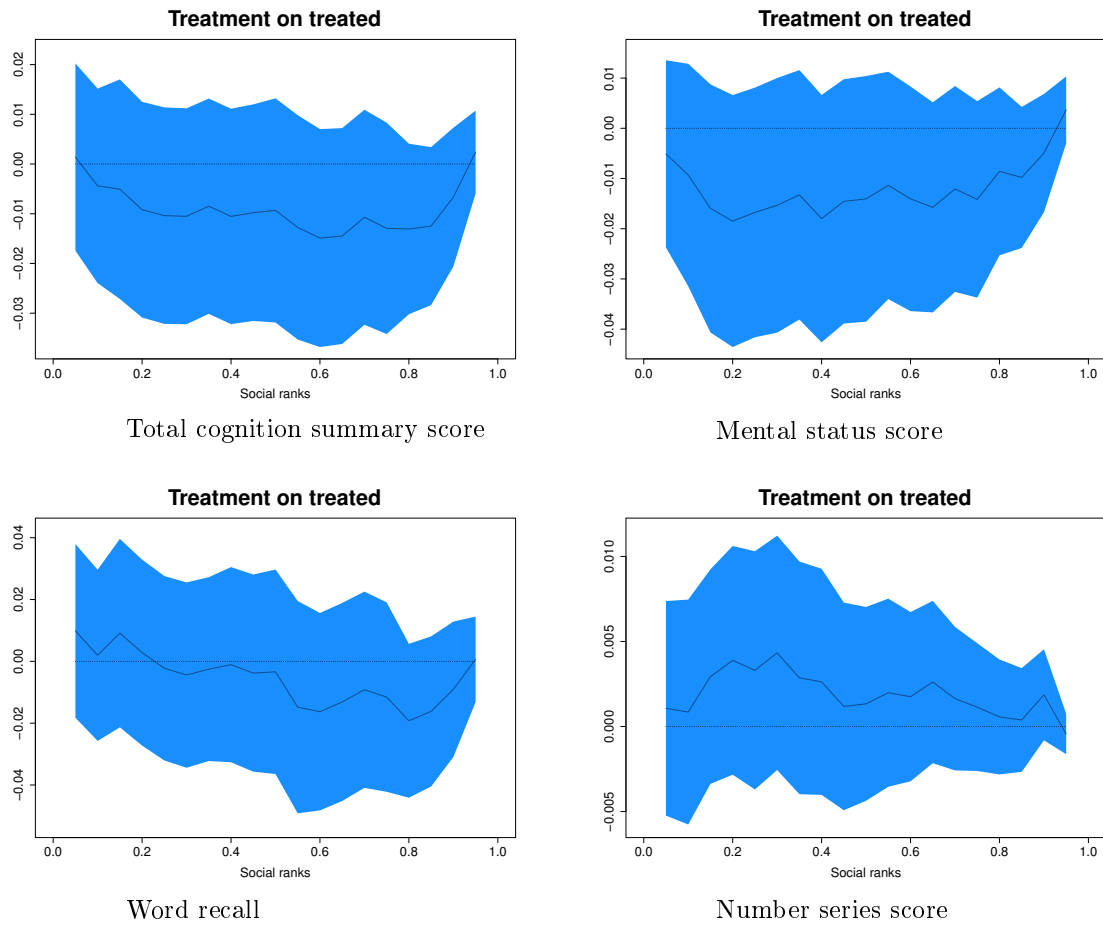
Note : The figure presents the treatment effect curves on the relative concentration curves of the four indicators, along with their 95% confidence intervals, estimated using the distributional DiD RIF regression. The sample is restricted to individuals in the control group (Jim Crow region), and the treatment is randomly assigned. The regression equations include covariates. For all four indicators, the treatment effect is not statistically significant.

FIGURE 2.24: Treatment effect on the treated for the absolute concentration curve-Placebo in the treatment group



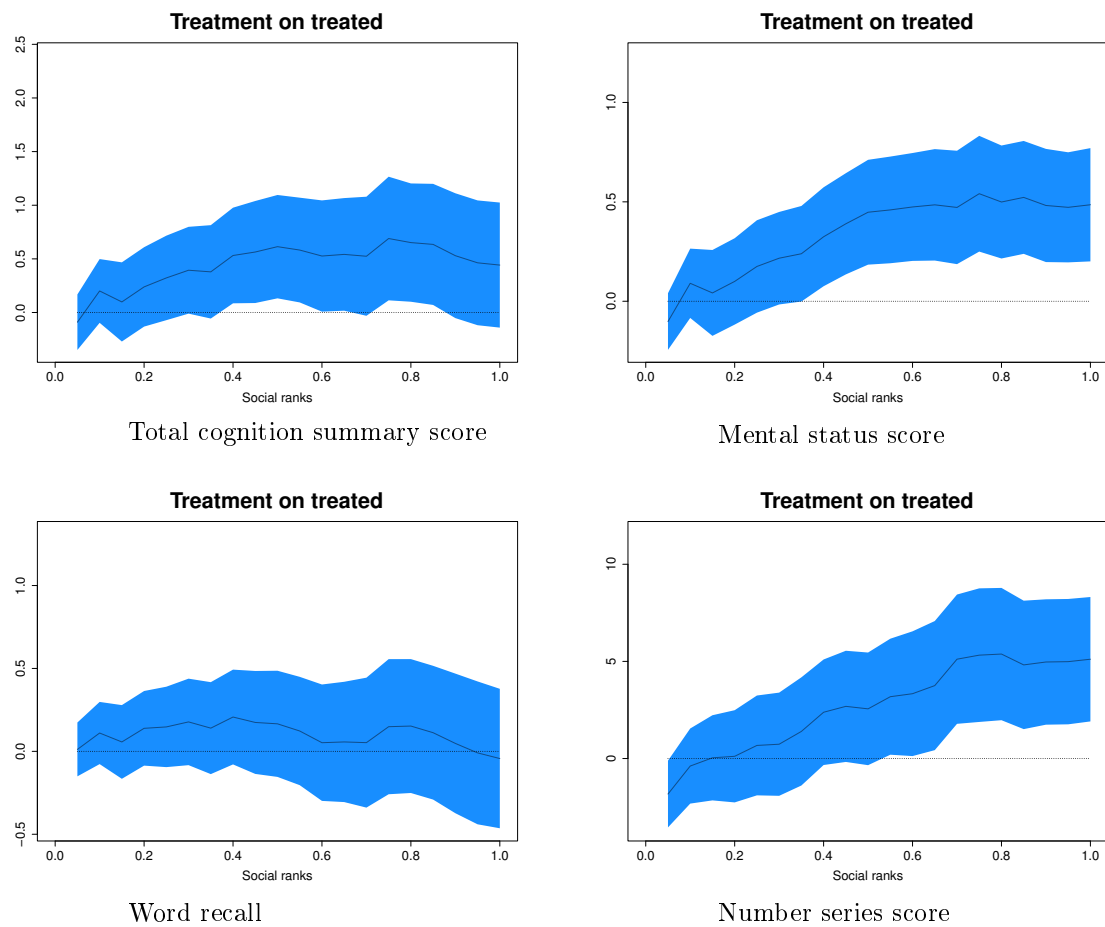
Note : The figure presents the treatment effect curves on the absolute concentration curves of the four indicators, along with their 95% confidence intervals, estimated using the distributional DiD RIF regression. The sample is restricted to individuals in the treatment group (Non-Jim Crow region), and the treatment is randomly assigned. The regression equations include covariates. For all four indicators, the treatment effect is not statistically significant.

FIGURE 2.25: Treatment effect on the treated for the relative concentration curve-Placebo in the treatment group



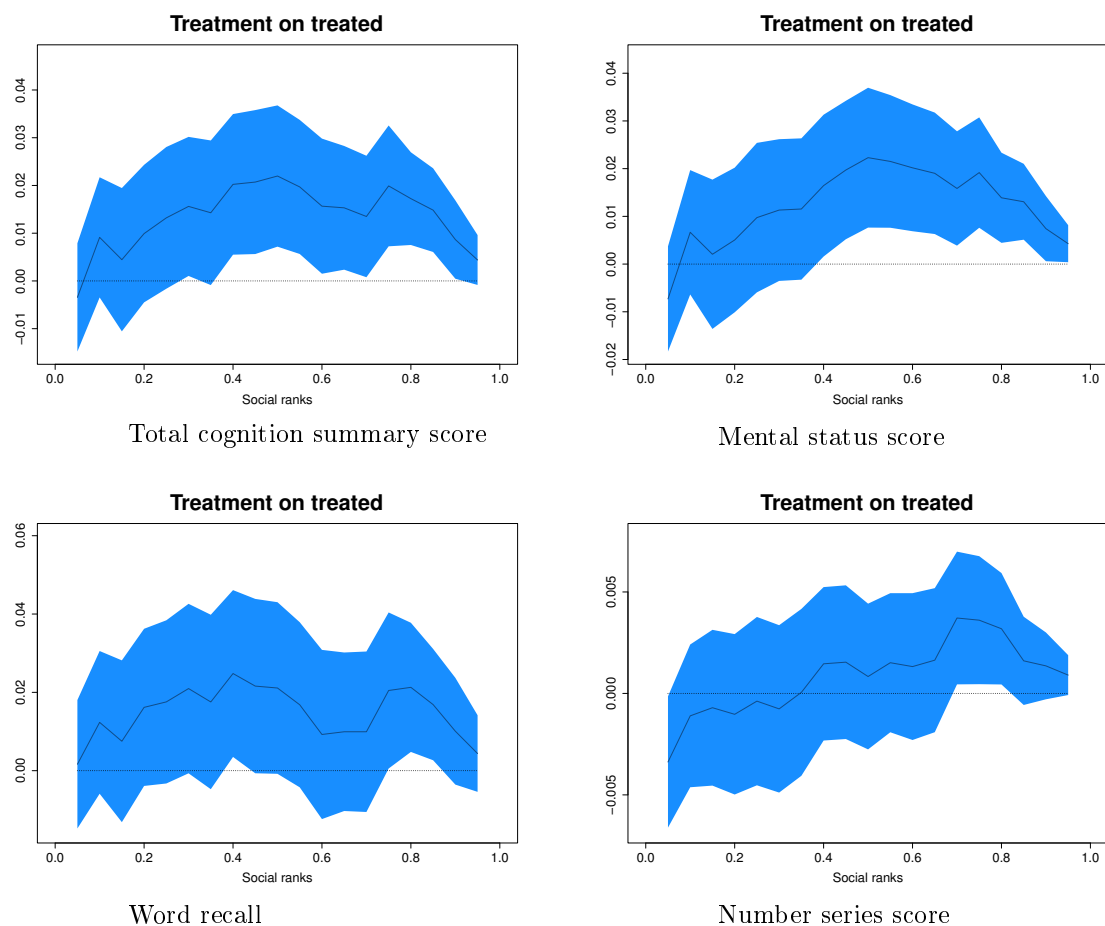
Note : The figure presents the treatment effect curves on the relative concentration curves of the four indicators, along with their 95% confidence intervals, estimated using the distributional DiD RIF regression. The sample is restricted to individuals in the treatment group (Non-Jim Crow region), and the treatment is randomly assigned. The regression equations include covariates. For all four indicators, the treatment effect is not statistically significant.

FIGURE 2.26: Treatment effect on the treated for the absolute concentration curve-Non-reweighted



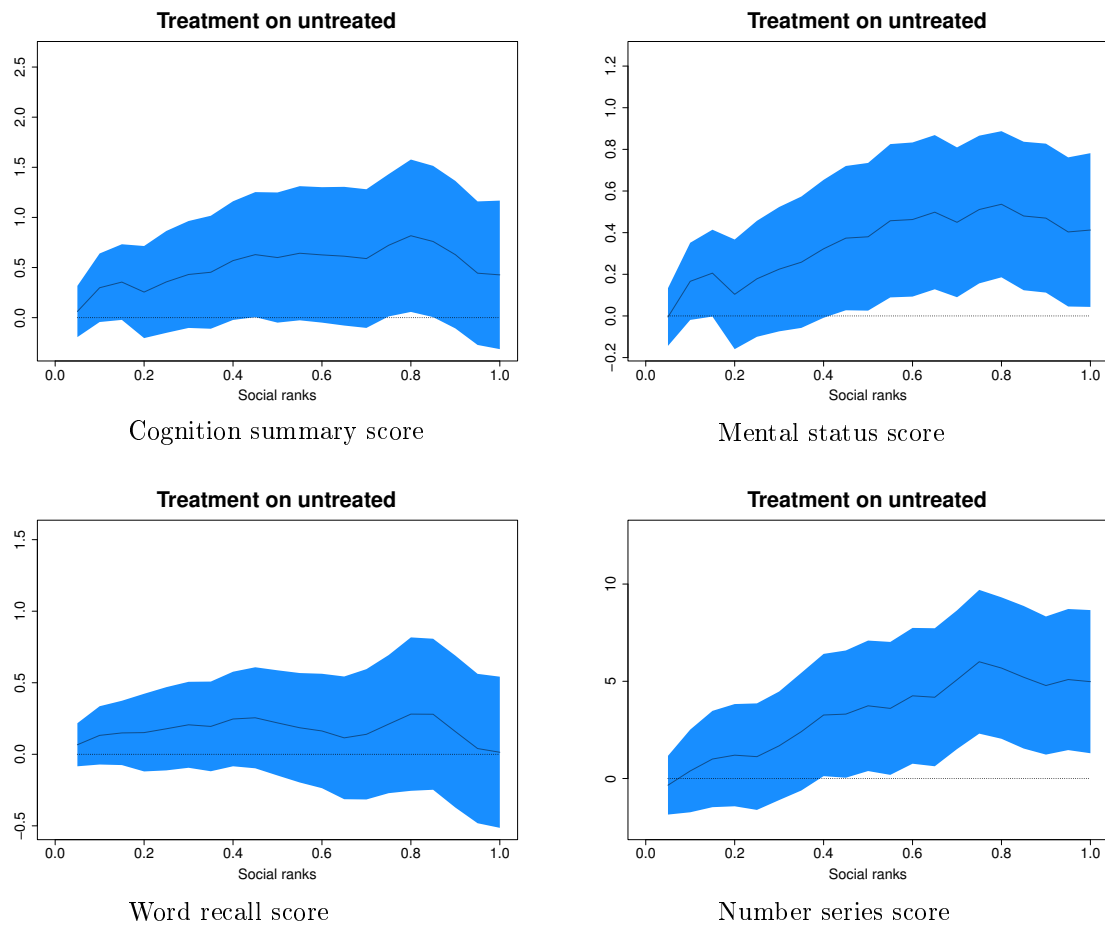
Note: This figure illustrates the estimated treatment effect curves on the absolute concentration curves of the four health indicators without reweighting the observations, as described in Subsection 2.5.2. Instead, survey weights are used. The regression equations include covariates. The figure shows robust results comparable to the original estimates.

FIGURE 2.27: Treatment effect on the treated for the relative concentration curve-Non reweighted



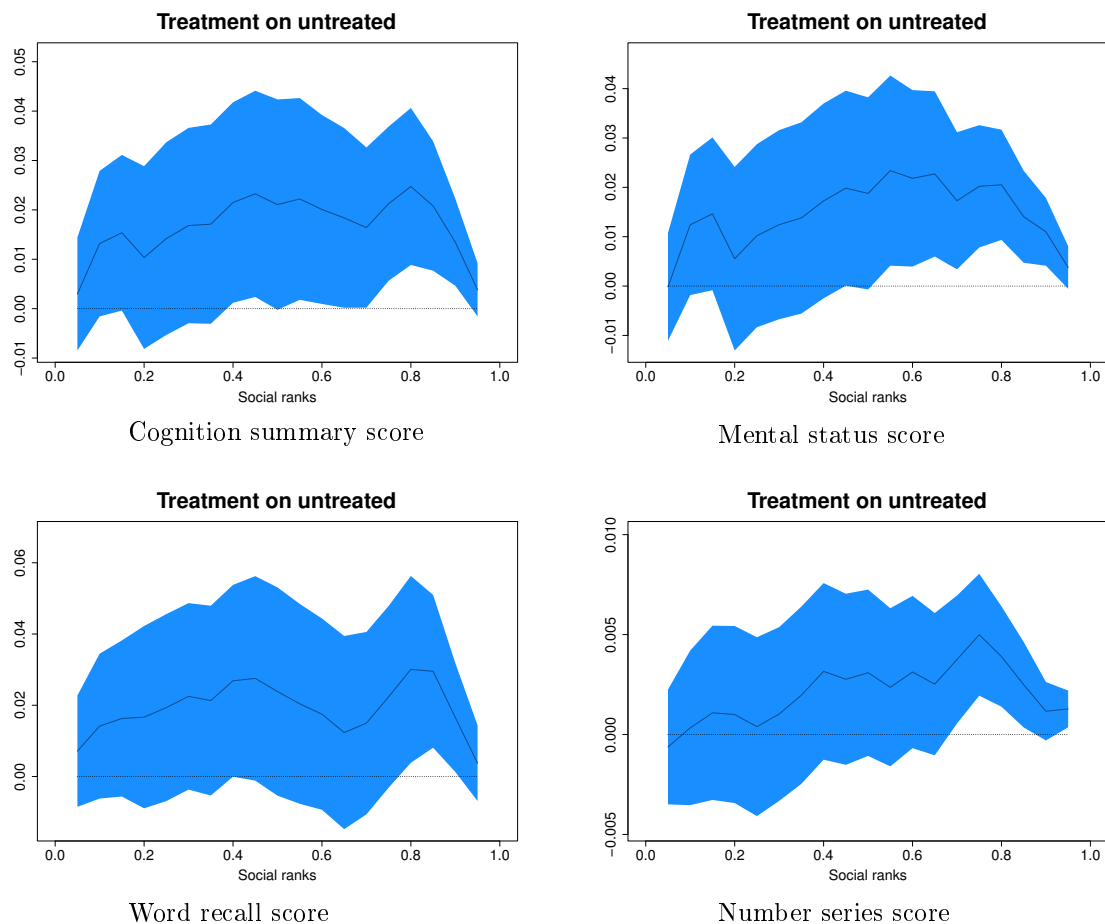
Note: This figure illustrates the estimated treatment effect curves on the relative concentration curves of the four health indicators without reweighting the observations, as described in Subsection 2.5.2. Instead, survey weights are used. The regression equations include covariates. The figure shows results that are robust and broadly comparable to the initial estimates, though somewhat less precise.

FIGURE 2.28: Treatment effect on the untreated for the absolute concentration curve



Note: This figure presents the estimated treatment effect curves on the untreated for the absolute concentration curves of the four health indicators. The regression equations include covariates. The results indicate a positive and statistically significant effect on the untreated for the mental status score and the number series score in the middle and upper parts of the distribution. In contrast, the effect is not significant for the word recall score and the cognition summary score, except at a few points in the distribution.

FIGURE 2.29: Treatment effect on the untreated for the relative concentration curve



Note: This figure presents the estimated treatment effect curves on the untreated for the relative concentration curves of the four health indicators. The regression equations include covariates. Regarding health inequality, the results indicate a positive, statistically significant effect of the treatment on the untreated, primarily in the upper part of the distribution.

Chapter 3

The dynamics of socioeconomic inequality of malnutrition in Egypt

3.1 Introduction

Malnutrition in Egypt presents a multifaceted challenge with significant repercussions for public health and economic vitality. It is estimated that malnutrition undercuts Egypt's annual GDP by nearly 2%, a testament to the severity of the issue (IDSC (Egyptian Cabinet Information and Decision Support Center), 2014). Moreover, the prevalence of malnutrition extends beyond the traditional specter of undernourishment, as evidenced by the disconcerting prevalence of overweight and obesity, particularly among women of reproductive age, where rates soar to 84% (Hu and Li, 2020). Concurrently, anemia exacerbates the health risks within this demographic, culminating in a dual health burden that impedes maternal and child health outcomes.

This chapter examines the changing patterns of malnutrition among Egyptian women across three time periods (2000, 2014, and 2021), with particular attention to underlying socioeconomic disparities. To measure these inequalities, we employ two complementary methodological approaches: the canonical health concentration index, which captures the socioeconomic inequalities in the distribution of health outcomes across socioeconomic ranks, and Wagstaff (2002) health shortfall index, which provides a more comprehensive assessment by combining population health averages with measures of socioeconomic inequality.

Building upon Abu-Ismaïl et al. (2020) work on child malnutrition in Egypt, we pivot to focus on the determinants of women’s malnutrition. Drawing from Heckley, Gerdtham, and Kjellsson (2016), we adopt a recentered influence function regression method to dissect the determinants of these health inequalities for both the health concentration and shortfall indices. This approach also informs our examination of the factors correlated to health shortfalls in this context.

The remainder of this chapter is structured as follows. Section 3.2 explains our methodological framework. Section 3.3 focuses on the empirical analysis of socioeconomic inequalities in overweightness, anemia, and the double burden of these two health outcomes. Section 3.4 concludes with a discussion of our findings and directions for future research.

3.2 Methodology

In this chapter, we quantify the prevalence of anemia and overweight among women of reproductive age, utilizing a continuous health variable, as advocated by Bilger, Kruger, and Finkelstein (2017). We identify anemia in non-pregnant women with a hemoglobin threshold below $12g/dL$ and in pregnant women below $11g/dL$. Accordingly, we define the health variable for anemia, a_i , as

$$a_i = \begin{cases} \max(0, 12 - h_i) & \text{if } i \text{ is not pregnant} \\ \max(0, 11 - h_i) & \text{if } i \text{ is pregnant,} \end{cases} \quad (3.1)$$

where h_i denotes the hemoglobin concentration for the i^{th} individual. An individual is overweighted when the body mass index (BMI) exceeds 25. Thus, our health variable for overweightness o_i , is

$$o_i = \max[0, bmi_i - 25], \quad (3.2)$$

where bmi_i represents the BMI for the i^{th} individual. To encapsulate the joint burden of anemia and overweight, we introduce the dual burden indicator:

$$d_i = \begin{cases} \max(0, 12 - h_i) \times I(bmi_i \geq 25) & \text{if } i \text{ is not pregnant} \\ \max(0, 11 - h_i) \times I(bmi_i \geq 25) & \text{if } i \text{ is pregnant,} \end{cases} \quad (3.3)$$

To measure socioeconomic health inequality, we rely on the health concentration index, the most widely used index in the literature on socioeconomic health inequality. For a joint distribution of health and income, F_{HY} , the health concentration index, $CI(F_{HY})$, is defined as

$$CI(F_{HY}) = \frac{2Cov(H, F_Y)}{\mu_H}, \quad (3.4)$$

where F_Y is the marginal distribution of income and μ_H is the average health status in the population.

Wagstaff (2002) argues that public health policymakers should usually focus simultaneously on the average health outcome and socioeconomic inequality in its distribution. He proposes an index, the health achievement index, that is a function of the average health status and these socioeconomic health inequalities. Since we are focusing on ill-health variables in this chapter, we use the term health shortfall index instead. The health shortfall index, $SI(F_{HY})$ is defined as

$$SI(F_{HY}) = \mu_H \times [1 - CI(F_{HY})], \quad (3.5)$$

The health shortfall index measures the severity of ill-health within a population, rising as the average ill-health outcome increases. It is also sensitive to changes in the distribution of health outcomes; it increases when the health concentration index diminishes. A declining health concentration index points to a greater concentration of the shift in health issues among the less affluent, suggesting that the change in adverse health outcomes disproportionately affects the poor. Thus, when the concentration index falls for a particular ill-health variable, it signals a widening socioeconomic inequality, with the burden increasingly borne by those in lower economic ranks.

It is also possible to rewrite the health shortfall index as a socially weighted average:

$$SI(F_{HY}) = \int_0^1 2(1-p)E[H|Y = Q - Y(p)], dp, \quad (3.6)$$

where $Q_Y(p)$ is the quantile function of income. This formulation of the health shortfall index encapsulates the social preferences that underlie the weighted average health status. The weighting function in this index is designed to reflect society's preference for the distribution of individual health statuses based on their income rank, indicating rank-dependent social preferences. Moreover, since the social weight function assigns greater weight to those with lower incomes, it puts more weight on the poorest segment of the population. For this reason, the social preferences underlying the construction of the health shortfall index display an aversion to socioeconomic health inequalities.

In this chapter, we also aim to examine the determinants of the health concentration and health shortfall indices for the three ill-health variables. This analysis estimates how a marginal change in the distribution of demographic characteristics may affect these indices. One way to identify these impacts is to use a recentered influence function regression (RIF-regression) approach (Firpo, Fortin, and Lemieux, 2009). The expected value of the recentered influence function provides a first-order von Mises approximation of a non-linear functional, such as the health concentration and health shortfall indices. Firpo, Fortin, and Lemieux (2009) show that the marginal effect of a non-linear functional is equal to the slope of the regression model of the individual values of the recentered influence function of the observations on the covariates of interest.

In this context, the analyst can use any standard regression approach to model $E[RIF|X]$ and use this model to evaluate these marginal effects.

Heckley, Gerdtham, and Kjellsson (2016) derived the expression of the recentered influence function of the health concentration index:

$$\widehat{RIF}(h_i, y_i; CI(F_{HY})) = -\frac{h_i}{\widehat{\mu}_H^2} 2\widehat{Cov}(H, F_Y) + \frac{1}{\widehat{\mu}_H} \left[\widehat{\mu}_H - h_i + 2h_i \widehat{F}_Y(y_i) - \frac{2}{N} \sum_{j=1}^N h_j \mathbb{I}(y_j \leq y_i) \right]. \quad (3.7)$$

Building on this result, Abu-Ismaïl et al. (2020) derived the expression of the recentered influence function of the health shortfall index:

$$\widehat{\text{RIF}}(h_i, y_i; SI(F_{HY})) = 2h_i(1 - p_i) - \widehat{\mu}_H + 2\widehat{\text{Cov}}(H, F_Y) + \frac{2}{N} \sum_{j=1}^N h_j \mathbb{I}(y_j \leq y_i). \quad (3.8)$$

We can econometrically model the health concentration and shortfall indices using these formulations. By regressing the estimated values of the recentered influence functions for each observation on covariates, we derive the marginal effects of changes in the covariate distribution on these indices.

3.3 Socioeconomic inequalities in anemia and overweightness in Egypt

3.3.1 The Egyptian context

In the evolving landscape of public health in the Arab Republic of Egypt, a paradoxical phenomenon of the malnutrition burden has emerged as a central concern, especially among women of reproductive age. Over the past three decades, Egypt has commendably improved its health indicators, with notable declines in maternal and infant mortality rates. Despite these strides, the nation faces a significant challenge: the coexistence of persistent undernutrition and escalating rates of overweight and obesity.

This dichotomy in nutritional health particularly impacts Egyptian women of reproductive age, where an alarming prevalence of overweightness and obesity intersects with persistent micronutrient deficiencies, notably anemia. This intersection of malnutrition forms a complex public health issue, with overweight women experiencing anemia at substantial rates. In their study of malnutrition in Egypt, Hu and Li (2020) estimate that anemia and obesity together affect 10.4% of Egyptian women of reproductive age in 2000. In 2014, this proportion increased to 11.3%. Such nutritional deficiencies are a marker of inadequate dietary quality and signal deeper systemic issues, including the impact of long-standing food subsidy policies that have historically favored energy-dense, nutrient-poor foods over nutrient-rich alternatives. Rising

food prices compel many families to opt for calorie-dense sustenance at the expense of balanced diets, exacerbating this situation.

The Ministry of Health and Population (MOHP) has recognized the gravity of this double burden and identified active steps to address malnutrition through comprehensive landscape analyses and the development of strategic nutrition policies (MOHP (Ministry of Health and Population), 2012). The report highlighted the need for a multifaceted approach to nutritional improvements, taking into account the socioeconomic and cultural factors that contribute to the malnutrition spectrum.

The analysis in this chapter offers one step in this needed multifaceted approach by focusing on the socioeconomic inequalities in anemia and overweight conditions and their intersection among Egyptian women of reproductive age. This demographic significantly influences the health trajectory of future generations. This focus is pertinent due to the direct health implications for women, the broader socioeconomic costs, and the potential for intergenerational perpetuation of malnutrition.

3.3.2 The data

This chapter analyzes the evolution of socioeconomic inequalities and nutritional deficiencies (i.e., shortfall) among women of reproductive age in Egypt by using data from the 2000 and 2014 Egyptian Demographic and Health Surveys (EDHS), complemented by the 2021 Egypt Family Health Survey. These nationally representative surveys provide comprehensive data on education, health, nutrition, and wealth. Our analysis focuses on nutritional outcomes – namely, anemia and overweightness – to investigate the dual burden of undernutrition and overnutrition jointly.

The Egypt Demographic and Health Survey (EDHS) relies on a nationally representative sampling frame based on the most recent population census conducted by the Central Agency for Public Mobilization and Statistics (CAPMAS). The survey adopts a multi-stage, stratified sampling design that covers all governorates of Egypt. Stratification is carried out by governorate and by urban–rural residence. Primary sampling units (PSUs) consist of shiakhnas or towns in urban areas and villages in rural areas,

from which households are selected in subsequent sampling stages. This design ensures adequate representation of the Egyptian population across geographic regions and settlement types.

Similarly, the 2021 Egypt Family Health Survey (EFHS) employs a multi-stage, stratified cluster sampling design to produce estimates that are representative at the national, regional, and governorate levels. The survey covers 26 governorates, excluding North Sinai, and is based on a sampling frame comprising 1,348 cadastral plots, from which approximately 25-35 households are selected per plot, yielding a total sample of about 34,000 households.

As with the EDHS, the sampling design and implementation of the EFHS are overseen by CAPMAS, ensuring consistency in sampling methodology and comparability across surveys. This common institutional framework strengthens the coherence of the two datasets and supports their joint use in empirical analysis. Table 3.1 compares the demographic characteristics of women across the two surveys. While some variation across years is observed, this is expected given temporal changes. Importantly, no sharp discontinuity appears between 2014 and 2021, despite the fact that these data come from two different surveys.

To evaluate nutritional status, we employ two primary health indicators: Body Mass Index (BMI) and hemoglobin concentration levels. To measure women's socioeconomic rank, it would ideally be appropriate to use individual or household income. However, neither survey provides information on income. Consequently, we use a wealth index as a proxy for income to determine individuals' relative positions within the wealth distribution. This index is measured at the household level and captures household assets, housing construction materials, and access to basic amenities such as water and sanitation facilities.

Our regression models incorporate personal characteristics, including age, total number of children born, current marital status, and the presence of a spouse, whether residing locally or employed overseas. For analyses of anemia and dual burden, we also include intrauterine contraceptive devices as an explanatory variable. All estimations use the survey-provided sample weights; however, heteroskedasticity is not explicitly

accounted for in the variance estimation.

Educational attainment is controlled for both women and their spouses, using “no formal education” as the baseline category, and distinguishing between “some primary education”, “primary to some secondary education”, and “secondary or higher education”. To address potential location effects, we designate the Cairo metropolitan governorate as the baseline category and include dummy variables for urban and rural areas in both Lower and Upper Egypt, as well as in the Frontier Governorates. These geographic controls allow us to capture the impact of regional attributes on the socioeconomic-related nutritional outcomes under study. Finally, housing quality indicators are not included as separate controls because they are already accounted for within the wealth index, which is a component of our dependent variable.

The initial sample comprised 15,573 observations for 2000, 21,762 for 2014, and 16,203 for 2021. To address missing data, we assume it is missing at random and exclude cases with missing data. As a result, the analytical dataset for overweightness includes 15,334 observations for 2000, 21,542 for 2014, and 16,087 for 2021.¹ Hemoglobin data, collected from random subsets, resulted in subsamples of 7,683 for 2000, 7,283 for 2014, and 9,525 for 2021. After applying the same assumption about missing data, the final sample sizes for hemoglobin analysis are 7,584 for 2000, 7,177 for 2014, and 9,475 for 2021. For dual burden analysis, which requires joint information on BMI and hemoglobin, the final sample sizes are 7,551 observations for 2000, 7,151 for 2014, and 9,432 for 2021.

Figure 3.1 illustrates the prevalence of overweightness, anemia, and the dual burden of malnutrition in Egypt from 2000 to 2021, highlighting significant trends in these three key health indicators. Specifically, overweightness exhibited a consistent upward trajectory, increasing from 75.4% [0.747, 0.761] in 2000 to 85.1% [0.845, 0.856] in 2021- a rise of 9.7 percentage points. In contrast, the prevalence of anemia exhibited a non-monotonic, initially declining from 27.8% [0.268, 0.288] in 2000 to 23.9% [0.229, 0.249] in 2014, before surging to 36.9% [0.360, 0.379] by 2021. The dual burden of malnutrition, defined as the coexistence of these conditions, displayed the most

¹Pregnant women could be excluded to account for pregnancy-related weight gain; nevertheless, the results remain largely unchanged.

striking trend: after remaining stable from 20.2% [0.193, 0.211] in 2000 to 19.0% [0.181, 0.199] in 2014, it increased sharply by 12.8 percentage points to 31.8% [0.309, 0.328] in 2021. These findings point to a deteriorating public health situation.

3.3.3 Malnutrition shortfall

In our empirical analysis, we first examine the evolution of the health shortfall index for the individual health indicators defined in Section 3.2. This provides a comprehensive assessment of malnutrition among Egyptian women of reproductive age. In the next subsection, we examine the health concentration index to quantify socioeconomic inequalities in malnutrition outcomes.

Figure 3.2 illustrates the trends of the shortfall index for anemia and the dual burden of malnutrition, while Figure 3.3 depicts the trajectory for the overweightness shortfall index. Tables 3.2, 3.3, and 3.4 present the results of the RIF-regressions of the health shortfall indices in the three malnutrition indicators.

The anemia shortfall index displays a complex trajectory over the study period, as shown in Figure 3.2. Starting at 0.342 in 2000 [0.320, 0.364], it decreased to 0.235 [0.215, 0.254] by 2014, before increasing sharply to 0.514 [0.485, 0.543] in 2021. The decline between 2000 and 2014 coincided with Egypt's national wheat flour fortification program (2008-2014), which aimed to combat micronutrient deficiencies by enriching wheat flour, a staple in "baladi" bread, widely consumed by lower-income households. However, the sharp increase observed in 2021 suggests that the earlier gains were not sustained, warranting further investigation.

As shown in Table 3.2, our RIF-regression analysis highlights evolving determinants of the anemia shortfall index over time. In 2000, age was the only significant demographic characteristic, exhibiting a small but statistically significant negative effect (p-value < 0.05).² This finding likely reflects biological and behavioral patterns: younger women, especially those in early menstruation stages, face higher anemia risks due to irregular and heavy menstrual cycles (Weyand et al., 2023). As women age, more

²The relationship between age and anemia is displayed in Figure 3.6. The figure suggests that a linear relationship between age and anemia can be reasonably assumed.

regular menstrual cycles and reduced pregnancy frequency may also help mitigate the risk of anemia associated with childbearing.

By 2014, the influence of the age effect diminished, and new significant factors emerged: number of children (positive effect), marital status (negative effect), and residence in Upper Egypt (positive effect). The negative association with marriage suggests potential protective factors: married women may benefit from better nutritional support and more comprehensive healthcare access, including regular prenatal and postnatal care, where anemia is routinely screened and treated. Conversely, the positive association with residence in Upper Egypt reflects the region's status as the country's poorest area, where limited access to iron-rich foods and healthcare services exacerbates anemia risks (Tang and Sholzberg, 2024).

The 2021 data introduces several notable changes. The use of an intrauterine contraceptive device emerges as a significant predictor effect at the 0.01 level. This strong association likely reflects the well-documented side effect of increased menstrual blood loss associated with intrauterine contraceptive device use, which can lead to or exacerbate iron deficiency anemia (Bathija et al., 1998; Liedholm et al., 1975). This biological mechanism helps explain the substantial coefficient and its high statistical significance. Education levels also become significant predictors, with both primary and secondary education showing positive associations. Simultaneously, the regional effects reverse direction, and all regions are now showing significant negative effects relative to urban governorates, suggesting a geographical restructuring of anemia risk.

Turning our attention to overnutrition as shown in Figure 3.3, we note that the overweightness shortfall index shows a consistent upward trend, rising from 4.083 [3.993, 4.173] in 2000 to 5.353 [5.259, 5.448] in 2014, and further increasing to 6.689 [6.573, 6.805] in 2021. The RIF-regression analyses presented in Table 3.3 reveal persistent and evolving socio-demographic patterns. Age maintained a significant positive impact across all three periods reflecting several mechanisms: the natural slowing of metabolism with age, leading to a reduced caloric expenditure; hormonal changes, particularly during menopausal transition, associated with increased abdominal fat; and greater social acceptance of larger body sizes among older women, particularly

after marriage or childbearing (MedlinePlus, 2024).³

Education shows an intriguing evolution. While having some education consistently exhibited positive effects on overweightness, by 2021, higher levels of education began showing non-significant coefficients, suggesting a potential for an educational gradient in overnutrition risk. This shift may indicate that growing health awareness among more educated women has started to translate into dietary changes and counteract the sedentary lifestyle that are often associated with education before 2021.

Regional disparities remain pronounced but evolved over time. The negative association between overweightness with residing in rural areas in Upper Egypt persisted across all periods, though its magnitude substantially decreased between 2000 and 2021. This persistence is likely due to the maintenance of more traditional, physically demanding agricultural lifestyles in these areas. Conversely, residing in rural areas in Lower Egypt shows an increasing positive association with being overweight, reflecting growing regional inequalities in overnutrition risk. This trend may result from differing rates of urbanization and lifestyle changes, with Lower Egypt experiencing more rapid transitions to sedentary lifestyles and greater access to energy-dense foods.

The dual burden shortfall index exhibits perhaps the most concerning trend, as shown in Figure 3.2. After remaining relatively stable—at 0.209 [0.192, 0.225] in 2000 and 0.188 [0.171, 0.206] in 2014, as indicated by overlapping confidence intervals—the index increased sharply to 0.435 [0.408, 0.462] by 2021. While Hu and Li (2020) employ a different methodology focusing exclusively on obesity ($\text{BMI} \geq 30$), they report only a slight increase in the dual burden between 2000 and 2014 (from 10.4% to 11.3%)⁴. In contrast, our index captures both overweight and obesity ($\text{BMI} \geq 25$) rather than obesity alone and incorporates gradations of anemia severity while accounting for its socioeconomic distribution. This more comprehensive measure, therefore, provides a more nuanced diagnostic tool for assessing Egypt's nutritional challenges.

The RIF-regression results presented in Table 3.4 underscore increasingly complex determinants of the dual burden. The use of intrauterine contraceptive devices (IUD)

³The relationship between age and overweightness is displayed in Figure 3.5. The figure suggests that a linear relationship between age and overweightness can be reasonably assumed.

⁴When considering obesity alone, the proportion of women experiencing the dual burden in our dataset is 10.6% in 2000, 10.5% in 2014, and 21.1% in 2021.

emerges as a consistently significant factor across all periods, with its coefficient increasing between 2000 and 2021. This persistent and strengthening association likely operates primarily through the anemia pathway, as IUD use is associated with increased menstrual blood loss, which can lead to or exacerbate iron deficiency anemia. The strengthening of this effect over time suggests that the impact of IUD-related blood loss may be particularly consequential in the context of concurrent overweightness. Beyond this consistent predictor, the 2021 results show significant positive associations with age, education, and number of children, while maintaining the negative association with marriage observed in 2014. The regional pattern shows a marked shift, with all regions displaying significant negative coefficients relative to urban governorates in 2021, contrasting with the mixed effects observed in earlier periods.⁵

These findings suggest that Egypt's nutritional challenges have intensified across all three indicators post-2014. The evolving pattern of socio-demographic determinants indicates the need for targeted interventions that account for changing risk factors. The persistence and strengthening of certain associations, particularly with intrauterine contraceptive device use, suggest specific areas where health policy interventions might be most effective. Moreover, shifting regional patterns indicate a need for geographically tailored approaches that account for local contexts and evolving risk factors.

3.3.4 Socioeconomic inequality in malnutrition

The concentration indices reveal striking temporal patterns in health inequalities (Figure 3.4). The overweightness concentration index shows a clear convergence toward equality, declining steadily from 0.135 [0.125, 0.145] in 2000 to 0.014 [0.006, 0.022] in 2014, and reaching -0.0003 [-0.009, 0.009] in 2021. This trend indicates that overweightness, as an ill-health condition, has evolved from being concentrated among the rich (pro-poor inequality) in 2000 to an almost neutral distribution across wealth quantiles by 2021, suggesting an equalization of overweightness risk across socioeconomic groups. The RIF-regression results presented in Table 3.6 highlight the changing nature of these disparities: while higher education levels showed significant negative

⁵The relationship between age and dual burden is displayed in Figure 3.7. The figure suggests that a linear relationship between age and dual burden can be reasonably assumed.

associations in 2000, they shifted to positive by 2021. Similarly, the strong positive effect of residing in rural Upper Egypt in 2000 diminishes substantially by 2021, indicating a geographical rebalancing of socioeconomic inequality in overweightness.

The anemia concentration index exhibits a notable reversal over time. Initially concentrated among the poor (pro-rich inequality) at -0.085 $[-0.119, -0.052]$ in 2000 and -0.104 $[-0.143, -0.064]$ in 2014, the anemia index becomes not significant at 0.011 $[-0.019, 0.041]$ in 2021. This shift suggests a fundamental restructuring of anemia risk across socioeconomic ranks. The RIF-regression results shown in Table 3.5 highlight key drivers of this shift. Regional effects, which were significant and positive in 2000 and 2014 –particularly in Lower Egypt– turned significantly negative by 2021, especially in rural areas. Additionally, the emergence of a husband’s primary education as a significant positive predictor of anemia risk in 2021 underscores shifting socioeconomic patterns influencing this health outcome. The disappearance of the previously significant protective effects of marriage by 2021 indicates evolving social determinants of inequality in anemia.

The dual burden concentration index demonstrates the most complex trajectory, moving from a statistically insignificant concentration at 0.019 $[-0.021, 0.058]$ in 2000 to a significant concentration among the poor (pro-rich inequality) at -0.081 $[-0.125, -0.037]$ in 2014, before returning to no significant socioeconomic inequality at 0.025 $[-0.008, 0.059]$ in 2021. While this equalization of socioeconomic inequality by 2021 might appear positive, it must be interpreted alongside the substantial increase in the health shortfall index from 0.188 $[0.171, 0.206]$ in 2014 to 0.435 $[0.408, 0.462]$ in 2021. This combination of patterns suggests a concerning trend: the dual burden of malnutrition has become more pervasive across all socioeconomic ranks, affecting a more significant proportion of the population more severely. The RIF-regression results presented in Table 3.7 reveal shifting regional patterns as key drivers: rural areas in Upper Egypt show a significant positive effect in 2000, becoming non-significant in 2014, and turning significantly negative by 2021. Similarly, Lower Egypt regions show significant negative effects by 2021, suggesting a geographical reorganization of the dual burden risk beyond socioeconomic status. Age effects also show interesting dynamics, shifting from positive in 2000 to negative in 2014, and becoming non-significant by 2021,

indicating changing demographic patterns in dual burden risk that mirror its wider population spread.

These evolving patterns indicate a complex transformation in Egypt's nutritional inequalities. The convergence in overweightness distribution across wealth ranks, combined with the reversal in anemia patterns and the fluctuating dual burden distribution, suggests that traditional socioeconomic gradients in malnutrition risk are being reshaped by broader societal changes. The shifting regional and educational patterns indicate the need for more nuanced policy approaches that account for these evolving risk distributions. The consistent pattern of changing regional effects across all three indicators by 2021 is particularly noteworthy. This suggests fundamental shifts in the geographic distribution of nutritional risks that warrant careful policy attention and further investigation into underlying causal mechanisms.

3.4 Conclusion

This chapter provides an empirical analysis of malnutrition shortfalls and socioeconomic inequalities among Egyptian women of reproductive age from 2000 to 2021. The findings reveal troubling shifts in nutritional patterns, particularly after 2014, with one of the most striking changes observed in the anemia shortfall index. While the period from 2000 to 2014 showed encouraging progress in reducing anemia, this positive trend reversed significantly by 2021. Specifically, the anemia shortfall index rose sharply from 0.235 [0.215, 0.254] in 2014 to 0.514 [0.485, 0.543] in 2021, with this deterioration affecting all socioeconomic groups equally, leading to the disappearance of socioeconomic inequality by 2021. Similarly, the overweightness shortfall index continued its upward trajectory, increasing from 5.353 [5.259, 5.448] in 2014 to 6.689 [6.573, 6.805] in 2021. This increase was accompanied by a convergence toward socioeconomic equality, underscoring the pervasive nature of this nutritional challenge across all economic groups.

One of the most concerning patterns is the evolution of the dual burden of malnutrition. After remaining relatively stable between 2000 and 2014, it increased dramatically by 2021, with the associated shortfall index more than doubling, rising from 0.188 [0.171, 0.206] to 0.435 [0.408, 0.462]. While the pronounced concentration of

the dual burden among the poor observed in 2014 disappears by 2021, this apparent equalization conceals a deeper issue: the dual burden has become so widespread that it now severely affects women across all socioeconomic ranks.

Building on these findings, the RIF-regression analysis sheds light on the evolving determinants underlying these trends. By 2021, the use of intrauterine contraceptive devices emerged as a significant predictor of both anemia and the dual burden, underscoring the intricate relationship between reproductive health choices and nutritional outcomes. Additionally, regional patterns of nutritional risk shifted markedly, with all regions showing significant effects compared with urban governorates. This indicates a fundamental geographical reorganization of nutritional risks, moving beyond the urban-rural divides and reflecting the widespread and multifaceted nature of malnutrition in Egypt.

These findings underscore the urgent need for renewed policy attention and a fundamental recalibration of public health strategies in Egypt. The widespread deterioration in nutritional status, despite reductions in socioeconomic inequalities, suggests that policies focused primarily on income protection may be insufficient. Addressing malnutrition effectively requires a broader perspective that accounts for Egypt-specific policy and institutional dynamics, including reforms to the food subsidy system, changes in the composition and quality of ration card commodities, and the central role of subsidized baladi bread in household diets. In particular, shifts in subsidy design that prioritize caloric adequacy over nutritional quality may have contributed to rising overweight and persistent micronutrient deficiencies.

Future research should prioritize identifying the drivers of the post-2014 deterioration, with particular attention to how recent policy changes—such as adjustments to ration card eligibility, modifications to the food basket composition, and constraints on preventive nutrition and primary healthcare services—may have exacerbated these trends. Such insights are essential for designing interventions that address malnutrition across all socioeconomic groups while targeting the specific vulnerabilities of at-risk populations. Ultimately, our findings highlight the complex and multifaceted nature of malnutrition in Egypt, where rising average shortfalls and persistent geographical disparities call for coordinated, evidence-based policy responses.

Appendix

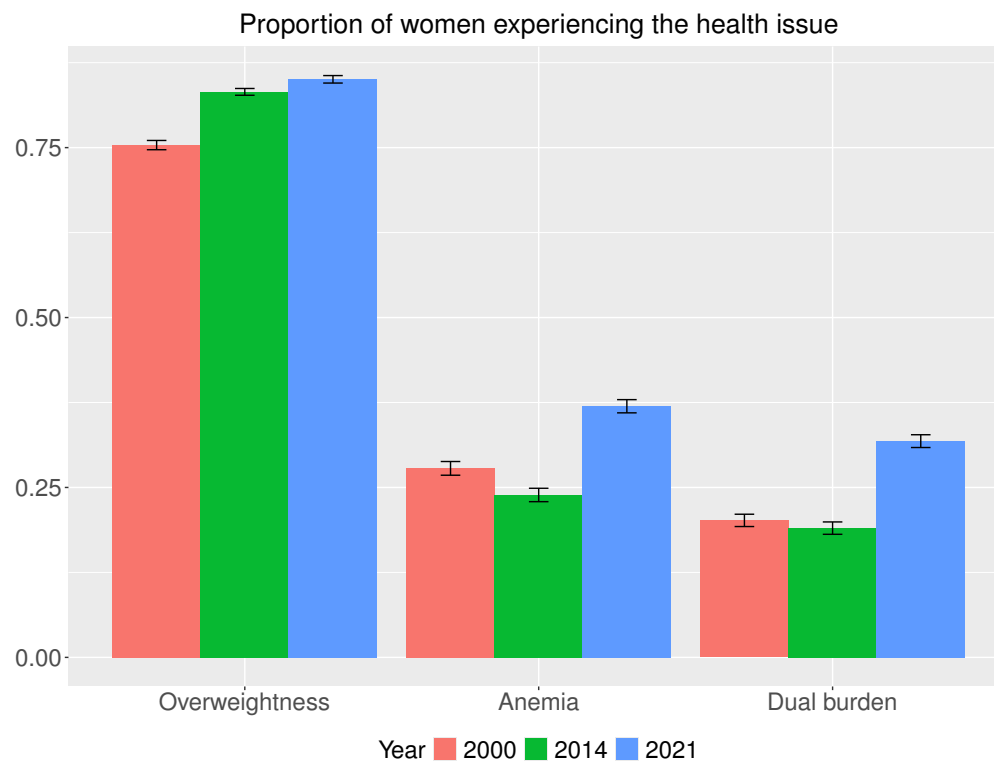
TABLE 3.1: Descriptive statistics by year

	2000	2014	2021
Respondent's current age	33.4	33.0	34.5
IUD used as contraception	32.8	28.3	27.4
The respondent is married	92.4	94.0	93.2
Husband absent from HH	3.8	4.6	5.1
Woman - Some primary	13.2	6.2	5.2
Woman - No education	43.4	24.0	16.7
Woman - Primary complete / some secondary	13.0	17.4	21.2
Woman - Secondary complete / higher	30.3	52.4	56.9
Husband - No education	30.1	17.1	13.4
Husband - Primary complete / some secondary	17.7	17.8	18.3
Husband - Secondary complete / higher	37.7	56.6	61.3
Urban governorates	19.1	12.7	15.2
Lower Egypt urban	12.5	10.6	11.2
Lower Egypt rural	31.5	38.4	32.3
Upper Egypt urban	11.5	11.1	12.3
Upper Egypt rural	24.1	26.2	28.1
Frontier governorates	1.3	0.9	0.9

Notes: With the exception of age, measured in years, all variables in the table are reported as proportions (%) of women with the corresponding characteristic.

Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021

FIGURE 3.1: Proportion of women with malnutrition deficiencies



Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021.

TABLE 3.2: Anemia shortfall index regressions

	2000	2014	2021
Age	-0.004** (0.002)	0.002 (0.001)	0.004 (0.002)
Some primary education	-0.007 (0.035)	-0.022 (0.038)	0.142** (0.067)
Some secondary education	0.031 (0.039)	0.075*** (0.029)	0.116** (0.045)
Secondary education or higher	-0.000 (0.039)	0.036 (0.026)	0.049 (0.042)
Number of children ever born	0.004 (0.006)	0.018*** (0.006)	0.008 (0.010)
Married	-0.043 (0.045)	-0.155*** (0.036)	-0.149*** (0.054)
Intrauterine contraceptive device (IUD)	0.044 (0.024)	0.024 (0.019)	0.188*** (0.030)
Husband absent	0.073 (0.060)	0.064 (0.040)	-0.013 (0.059)
Husband with some primary education	-0.011 (0.036)	-0.037 (0.036)	0.050 (0.060)
Husband with some secondary education	-0.078** (0.035)	-0.017 (0.030)	0.063 (0.048)
Husband with secondary education or higher	-0.068 (0.037)	-0.017 (0.028)	0.044 (0.044)
Lower Egypt urban	-0.029 (0.040)	0.046 (0.035)	-0.220*** (0.051)
Lower Egypt rural	-0.018 (0.033)	0.053 (0.028)	-0.306*** (0.041)
Upper Egypt urban	-0.041 (0.041)	0.060 (0.034)	-0.127** (0.051)
Upper Egypt rural	0.064 (0.037)	0.140*** (0.031)	-0.121*** (0.044)
Frontier governorate	-0.022 (0.098)	-0.003 (0.088)	-0.253** (0.127)
Intercept	0.541*** (0.081)	0.192*** (0.066)	0.527*** (0.103)
R^2	0.005	0.011	0.014
N	7,584	7,175	9,475

Notes: Standard errors in parentheses.

***: p-value < 0.01; **: p-value < 0.05

Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021

TABLE 3.3: Overweightness shortfall index regressions

	2000	2014	2021
Age	0.175*** (0.006)	0.166*** (0.006)	0.197*** (0.008)
Some primary education	0.755*** (0.126)	0.206 (0.173)	1.245*** (0.256)
Some secondary education	0.807*** (0.136)	0.375*** (0.129)	0.508*** (0.178)
Secondary education or higher	0.611*** (0.137)	0.129 (0.120)	-0.110 (0.166)
Number of children ever born	0.015 (0.022)	0.251*** (0.028)	0.243*** (0.039)
Married	0.360** (0.153)	0.540*** (0.164)	0.220 (0.205)
Husband absent	0.420** (0.206)	-0.250 (0.182)	0.004 (0.230)
Husband with some primary education	0.282** (0.128)	0.017 (0.161)	0.779*** (0.238)
Husband with some secondary education	0.475*** (0.125)	-0.145 (0.134)	0.208 (0.188)
Husband with secondary education or higher	0.278** (0.131)	-0.097 (0.125)	-0.096 (0.173)
Lower Egypt urban	0.806*** (0.142)	0.497*** (0.157)	0.796*** (0.198)
Lower Egypt rural	0.339*** (0.119)	1.111*** (0.124)	1.116*** (0.157)
Upper Egypt urban	-0.764*** (0.146)	-0.162 (0.156)	0.011 (0.193)
Upper Egypt rural	-3.136*** (0.131)	-1.515*** (0.137)	-1.819*** (0.166)
Frontier governorate	-1.920*** (0.347)	-0.904** (0.416)	-1.060** (0.533)
Intercept	-2.763*** (0.283)	-1.501*** (0.302)	-1.082*** (0.399)
R^2	0.178	0.113	0.105
N	15,353	21,518	16,087

Notes: Standard errors in parentheses.

***: p-value < 0.01; **: p-value < 0.05

Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021

TABLE 3.4: Dual burden shortfall index regressions

	2000	2014	2021
Age	0.001 (0.001)	0.004*** (0.001)	0.007*** (0.002)
Some primary education	-0.007 (0.026)	0.004 (0.034)	0.134** (0.063)
Some secondary education	0.013 (0.029)	0.085*** (0.026)	0.067 (0.042)
Secondary education or higher	0.004 (0.029)	0.047 (0.024)	0.025 (0.040)
Number of children ever born	0.002 (0.005)	0.012** (0.006)	0.025*** (0.010)
Married	-0.037 (0.033)	-0.085*** (0.033)	-0.157*** (0.050)
Intrauterine contraceptive device (IUD)	0.050*** (0.018)	0.035** (0.017)	0.166*** (0.028)
Husband absent	0.034 (0.045)	0.035 (0.037)	0.030 (0.055)
Husband with some primary education	-0.018 (0.027)	-0.019 (0.032)	0.096 (0.056)
Husband with some secondary education	-0.030 (0.026)	-0.027 (0.027)	0.098** (0.045)
Husband with secondary education or higher	-0.029 (0.027)	-0.026 (0.025)	0.086** (0.042)
Lower Egypt urban	-0.015 (0.030)	0.046 (0.032)	-0.162*** (0.048)
Lower Egypt rural	-0.015 (0.025)	0.055** (0.025)	-0.226*** (0.038)
Upper Egypt urban	-0.056 (0.030)	0.043 (0.031)	-0.098** (0.048)
Upper Egypt rural	-0.088*** (0.028)	0.081*** (0.028)	-0.154*** (0.041)
Frontier governorate	-0.038 (0.073)	-0.035 (0.080)	-0.218 (0.119)
Intercept	0.250*** (0.060)	0.027 (0.060)	0.259*** (0.096)
R^2	0.004	0.009	0.015
N	7,551	7,149	9,432

Notes: Standard errors in parentheses.

***: p-value < 0.01; **: p-value < 0.05

Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021

TABLE 3.5: Anemia concentration index regressions

	2000	2014	2021
Age	0.002 (0.003)	-0.004 (0.003)	-0.002 (0.002)
Some primary education	0.032 (0.052)	0.111 (0.079)	-0.114 (0.067)
Some secondary education	0.061 (0.057)	-0.052 (0.059)	-0.020 (0.045)
Secondary education or higher	-0.054 (0.057)	-0.031 (0.055)	-0.011 (0.042)
Number of children ever born	-0.003 (0.009)	-0.021 (0.013)	0.020 (0.010)
Married	-0.056 (0.066)	0.231*** (0.075)	-0.018 (0.054)
Intrauterine contraceptive device (IUD)	0.045 (0.036)	0.051 (0.040)	-0.017 (0.030)
Husband absent	-0.023 (0.088)	-0.219*** (0.084)	0.047 (0.059)
Husband with some primary education	0.027 (0.053)	0.095 (0.074)	0.158*** (0.060)
Husband with some secondary education	0.090 (0.052)	0.054 (0.062)	-0.011 (0.048)
Husband with secondary education or higher	-0.031 (0.054)	-0.012 (0.058)	0.015 (0.044)
Lower Egypt urban	0.121** (0.059)	0.100 (0.073)	-0.132*** (0.051)
Lower Egypt rural	0.150*** (0.049)	0.066 (0.058)	-0.070 (0.041)
Upper Egypt urban	0.082 (0.060)	0.188*** (0.071)	-0.082 (0.051)
Upper Egypt rural	0.045 (0.055)	-0.014 (0.064)	-0.167*** (0.044)
Frontier governorate	0.197 (0.144)	0.352 (0.184)	-0.140 (0.127)
Intercept	-0.267** (0.119)	-0.278** (0.138)	0.142 (0.103)
R^2	0.004	0.007	0.003
N	7,584	7,175	9,475

Notes: Standard errors in parentheses.

***: p-value < 0.01; **: p-value < 0.05

Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021

TABLE 3.6: Overweightness concentration index regressions

	2000	2014	2021
Age	-0.003*** (0.001)	0.001 (0.001)	0.0004 (0.001)
Some primary education	-0.029** (0.014)	0.028 (0.016)	0.005 (0.020)
Some secondary education	-0.006 (0.016)	0.026** (0.012)	0.041*** (0.014)
Secondary education or higher	-0.118*** (0.016)	0.012 (0.011)	0.036*** (0.013)
Number of children ever born	-0.010*** (0.002)	-0.006** (0.003)	0.004 (0.003)
Married	0.025 (0.018)	-0.011 (0.015)	-0.039** (0.016)
Husband absent	-0.033 (0.024)	0.037** (0.017)	0.043** (0.018)
Husband with some primary education	0.002 (0.015)	0.029** (0.015)	0.004 (0.018)
Husband with some secondary education	0.003 (0.014)	0.054*** (0.012)	0.018 (0.015)
Husband with secondary education or higher	0.009 (0.015)	0.008 (0.012)	0.003 (0.013)
Lower Egypt urban	0.003 (0.016)	0.083*** (0.015)	-0.008 (0.015)
Lower Egypt rural	-0.084*** (0.014)	0.015 (0.012)	-0.079*** (0.012)
Upper Egypt urban	0.011 (0.017)	0.051*** (0.014)	-0.017 (0.015)
Upper Egypt rural	0.221*** (0.015)	0.178*** (0.013)	0.065*** (0.013)
Frontier governorate	0.123*** (0.040)	0.087** (0.039)	0.012 (0.041)
Intercept	0.383*** (0.032)	-0.073*** (0.028)	-0.018 (0.031)
R^2	0.055	0.019	0.014
N	15,353	21,518	16,087

Notes: Standard errors in parentheses.

***: p-value < 0.01; **: p-value < 0.05

Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021

TABLE 3.7: Dual burden concentration index regressions

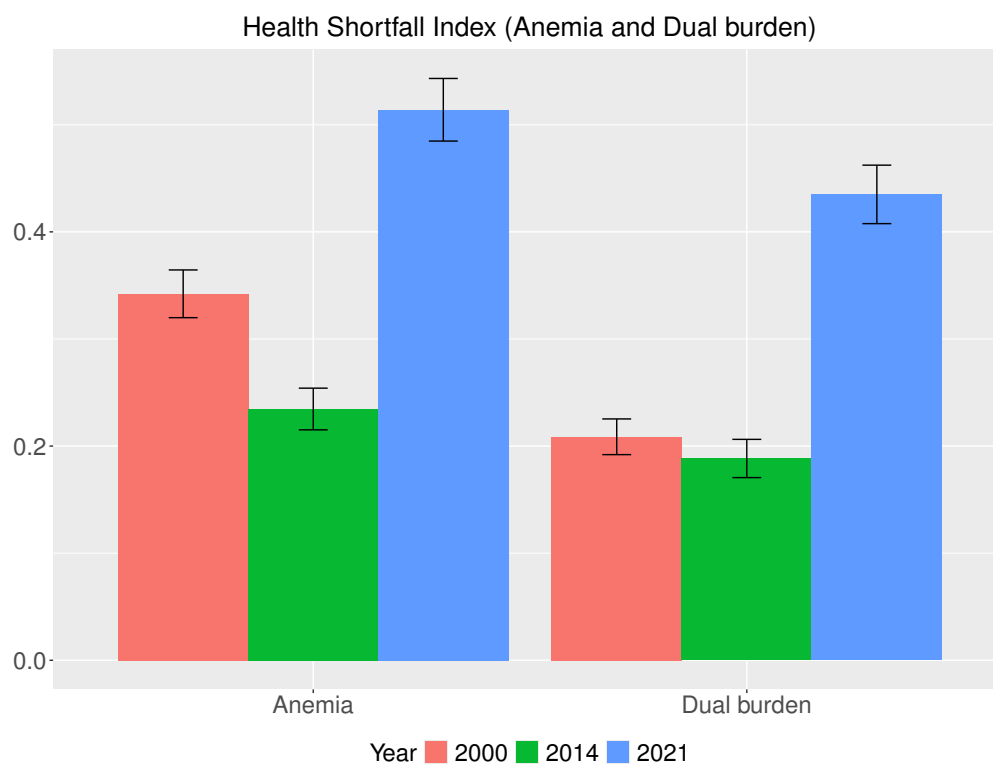
	2000	2014	2021
Age	0.006** (0.003)	-0.006** (0.003)	-0.003 (0.002)
Some primary education	0.076 (0.061)	0.101 (0.088)	-0.092 (0.074)
Some secondary education	0.107 (0.067)	-0.106 (0.066)	0.033 (0.050)
Secondary education or higher	-0.057 (0.067)	-0.059 (0.061)	0.026 (0.047)
Number of children ever born	-0.018 (0.011)	-0.008 (0.015)	0.017 (0.011)
Married	0.053 (0.077)	0.128 (0.084)	0.003 (0.059)
Intrauterine contraceptive device (IUD)	0.035 (0.042)	0.066 (0.045)	-0.033 (0.033)
Husband absent	0.050 (0.104)	-0.143 (0.093)	0.003 (0.065)
Husband with some primary education	0.075 (0.062)	0.114 (0.082)	0.100 (0.067)
Husband with some secondary education	0.058 (0.060)	0.118 (0.069)	-0.067 (0.053)
Husband with secondary education or higher	-0.046 (0.064)	0.033 (0.065)	-0.041 (0.049)
Lower Egypt urban	0.094 (0.069)	0.118 (0.081)	-0.127** (0.057)
Lower Egypt rural	0.109 (0.058)	0.049 (0.064)	-0.109** (0.045)
Upper Egypt urban	0.060 (0.071)	0.174** (0.079)	-0.105 (0.057)
Upper Egypt rural	0.232*** (0.064)	0.032 (0.071)	-0.159*** (0.048)
Frontier governorate	0.241 (0.170)	0.417** (0.205)	-0.156 (0.141)
Intercept	-0.291** (0.139)	-0.137 (0.153)	0.235** (0.114)
R^2	0.005	0.005	0.003
N	7,551	7,149	9,432

Notes: Standard errors in parentheses.

***: p-value < 0.01; **: p-value < 0.05

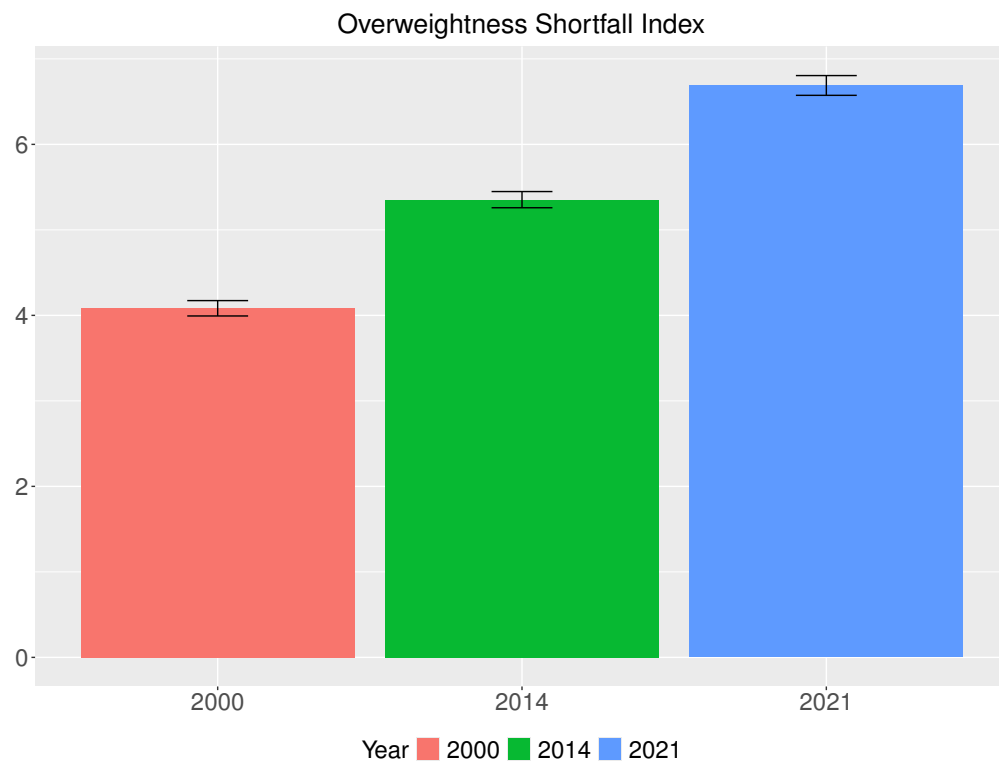
Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021

FIGURE 3.2: Shortfall in anemia and dual burden of anemia and overweightness



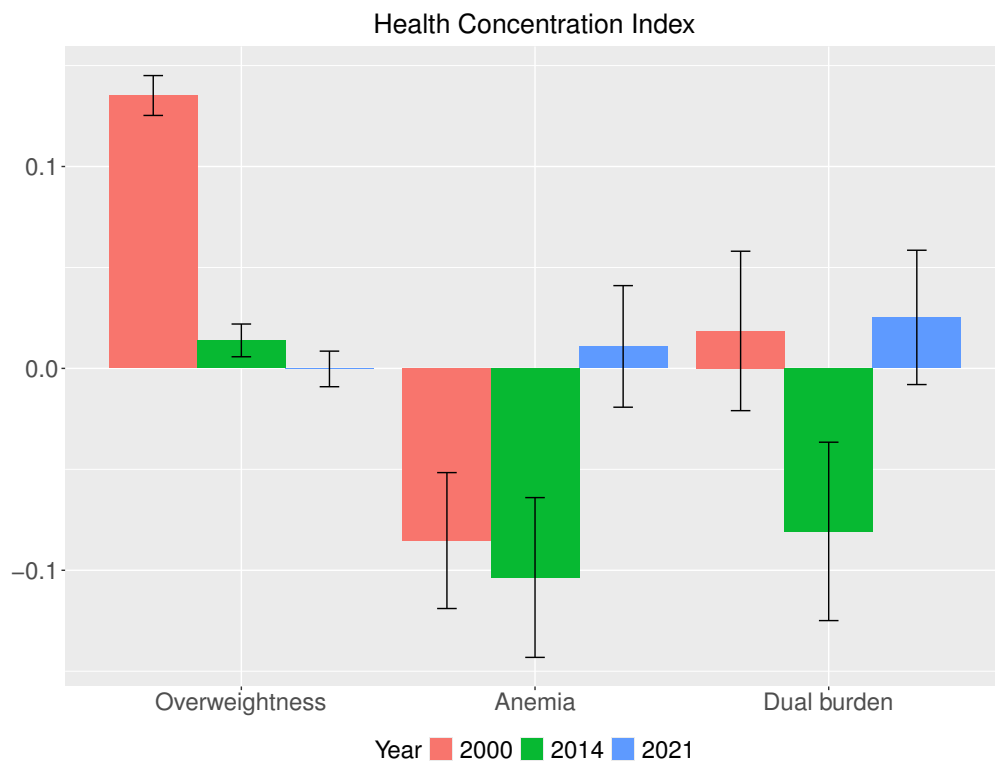
Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021.

FIGURE 3.3: Shortfall in overweightness



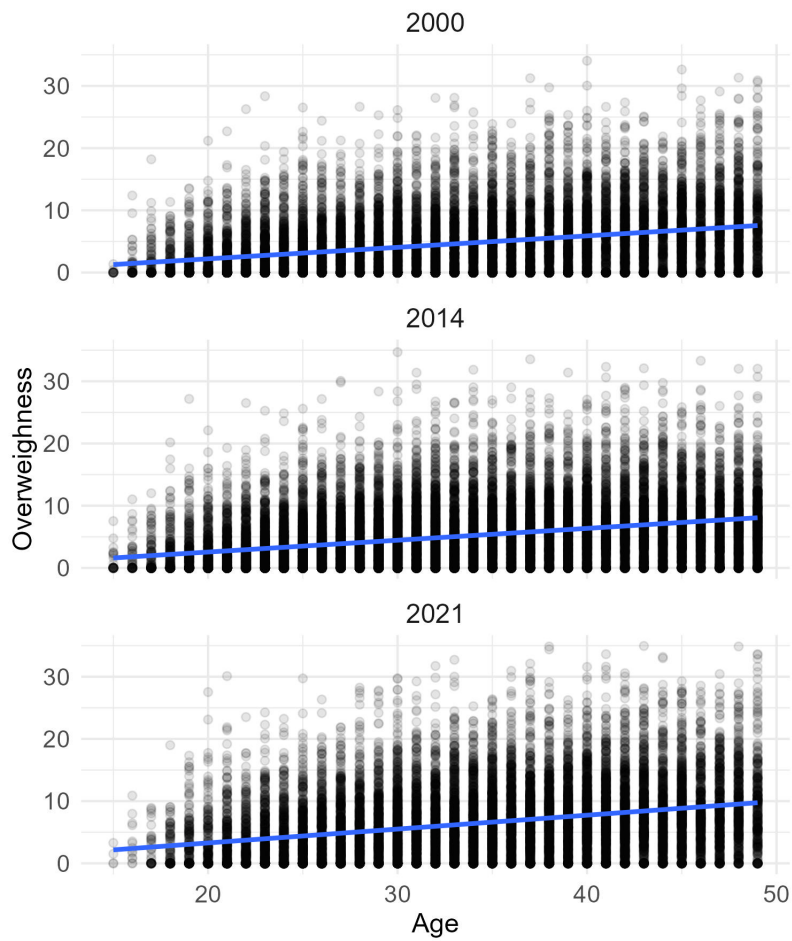
Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021.

FIGURE 3.4: Socioeconomic inequalities in malnutrition



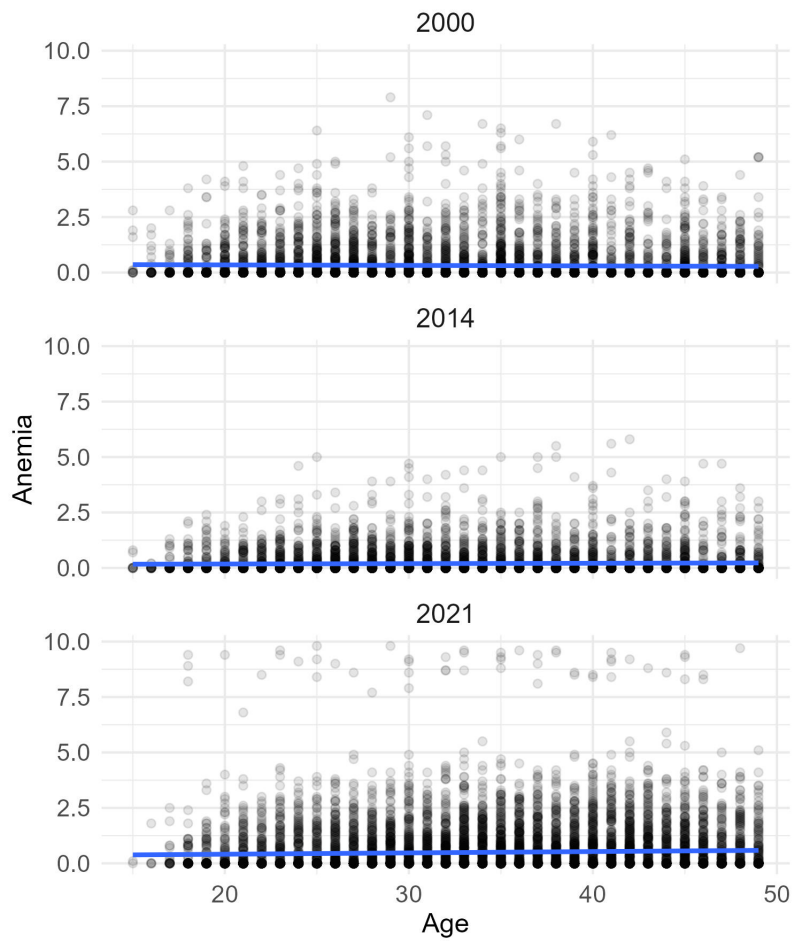
Source: Authors' own estimations, EDHS 2000, EDHS 2014, & EFHS 2021.

FIGURE 3.5: Overweightness by age



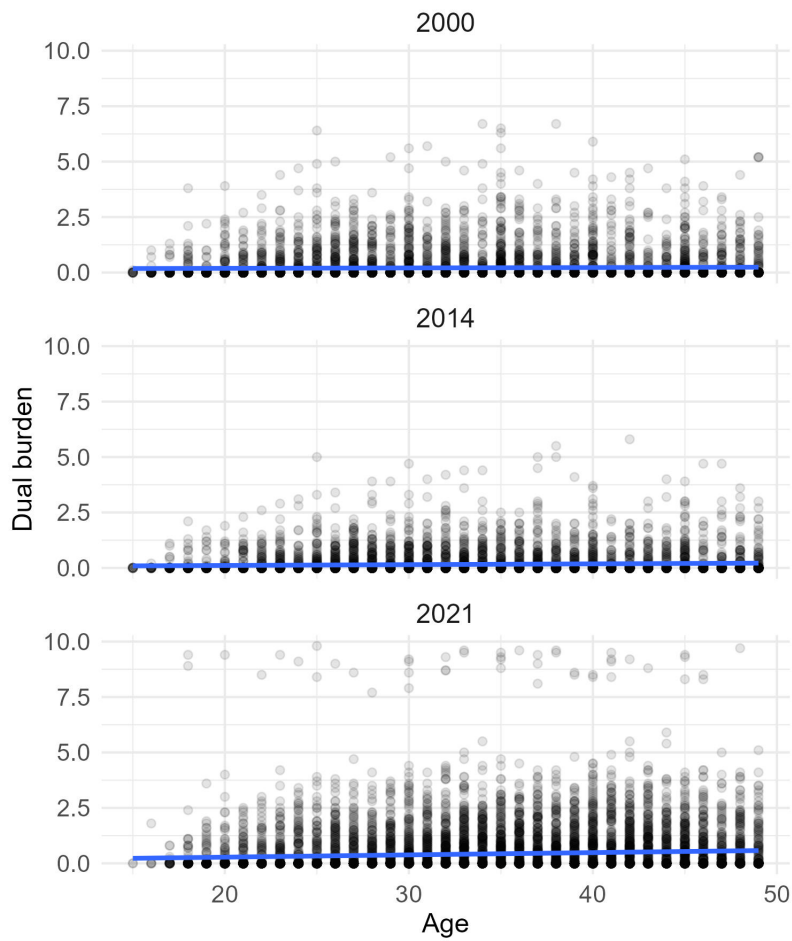
The figure displays the relationship between age and overweightness. Each point represents an individual observation, while the blue line shows the estimated linear regression of overweightness on age. The panels correspond to different survey years.

FIGURE 3.6: Anemia by age



The figure displays the relationship between age and anemia. Each point represents an individual observation, while the blue line shows the estimated linear regression of anemia on age. The panels correspond to different survey years.

FIGURE 3.7: Dual burden by age



The figure displays the relationship between age and dual burden. Each point represents an individual observation, while the blue line shows the estimated linear regression of dual burden on age. The panels correspond to different survey years.

General Conclusion

This thesis advances our understanding of health inequalities by combining empirical and methodological contributions. Across the three chapters, we examine the persistence, evolution, and determinants of health disparities, while proposing a distributional methodology to better measure and interpret them.

In Chapter 1, we adapt the RIF proposed by Firpo, Fortin, and Lemieux (2009) to the absolute and the relative concentration curves and derive explicit expressions for the influence function and the RIF for both curves, providing tools that can be readily implemented with standard econometric tools. In addition, we adapt the reweighting approach of DiNardo, Fortin, and Lemieux (1996) to construct counterfactual absolute and relative concentration curves. Thus, we apply these two approaches to decompose the change in these two curves between 2020 and 2000. Applying this framework to BMI and smoking data from the NHIS in the United States, we show how health inequalities are distributed across the socioeconomic spectrum. The reweighting method alone enables a global decomposition, while the RIF-based approach allows for isolating the contribution of individual determinants.

The second chapter examines the long-term effects of legalized segregation under Jim Crow laws on the cognitive health of older Black Americans. Employing a distributional difference-in-differences framework combined with the RIF regression, we move beyond average treatment effects to evaluate how structural policies influenced health outcomes across the socioeconomic ranks. Our findings reveal that the absence of Jim Crow laws conferred significant cognitive benefits, particularly for lower- and middle-income distribution, reducing health disparities and highlighting the enduring impact of legalized segregation on older adults' health.

The third chapter investigates trends in malnutrition among Egyptian women of reproductive age from 2000 to 2021. While the results show reductions in anemia between 2000 and 2014, subsequent deterioration by 2021 affected all socioeconomic groups, effectively erasing inequality but reflecting a widespread public health challenge. Overweightness and the dual burden of malnutrition also intensified, becoming increasingly pervasive. RIF-regression analysis identifies evolving determinants, including contraceptive use and geographic disparities, illustrating the multifactorial and context-dependent nature of nutritional inequalities. These findings underscore the limitations of interventions that focus solely on economic disparities and call for comprehensive strategies to address malnutrition.

In sum, the three chapters underscore the importance of understanding the determinants of socioeconomic health inequalities and health achievement/shortfall. They contribute to the literature by offering three distinct applications that integrate methodological approaches with empirical analysis. These contributions quantify the extent of health inequality and health achievement/shortfall, and identify some of its underlying drivers, thereby providing valuable insights for both academic research and policy design.

Bibliography

- Abu-Ismaïl, K., V. Gantner, P. Makdissi, and M. Yazbeck (2020). “Socioeconomic inequalities in child malnutrition in Egypt”. *Metron* 78.2, pp. 175–191.
- Alexander, J. Trent, Christine Leibbrand, Catherine Massey, and Stewart Tolnay (2017). “Second-Generation Outcomes of the Great Migration”. *Demography* 54.6, pp. 2249–2271. DOI: [10.1007/s13524-017-0619-9](https://doi.org/10.1007/s13524-017-0619-9).
- Anderson, K. F. (2016). “Residential Segregation and Health Outcomes in the United States: Moving Beyond Black and White”. *Population Review* 55.2. DOI: [10.1353/prv.2016.0006](https://doi.org/10.1353/prv.2016.0006).
- Austin, N., S. Harper, and E. Strumpf (2016). “Does Segregation Lead to Lower Birth Weight?: An Instrumental Variable Approach”. *Epidemiology* 27.5, 682–689. URL: <https://www.jstor.org/stable/26511859>.
- Bailey, Z. D., N. Krieger, M. Agénor, J. Graves, N. Linos, and M. T. Bassett (2017). “Structural racism and health inequities in the USA: evidence and interventions”. *The Lancet* 389.10077, 1453–1463.
- Bathija, H. et al. (Nov. 1998). “Effects of contraceptives on hemoglobin and ferritin”. English (US). *Contraception* 58.5, pp. 261–273. ISSN: 0010-7824. DOI: [10.1016/S0010-7824\(98\)00108-5](https://doi.org/10.1016/S0010-7824(98)00108-5).
- Bchi, K., P. Makdissi, and M. Yazbeck (2024). *A Modeling Approach to Decomposing Changes in Health Concentration Curves*. Working Paper 2403E. University of Ottawa, Department of Economics. URL: <http://hdl.handle.net/10393/46239>.
- Bilger, M., E.J. Kruger, and E.A. Finkelstein (2017). “Measuring socioeconomic inequality in obesity, looking beyond the obesity threshold”. *Health Economics* 26, pp. 1052–1056.
- Blinder, A. (1973). “Wage discrimination: reduced form and structural estimates”. *Journal of Human Resources* 8, pp. 436–455.

- Bondurant, S. and R. Wedge (2009). *Combating Tobacco Use in Military and Veteran Populations*. Ed. by Institute of Medicine (US) Committee on Smoking Cessation in Military and Veteran Populations. Accessed: 2026-01-31. Washington (DC): National Academies Press (US). Chap. 3, FACTORS THAT INFLUENCE TOBACCO USE. URL: <https://www.ncbi.nlm.nih.gov/books/NBK215329/>.
- Caunca, M. R., M. C. Odden, M. M. Glymour, T. Elfassy, K. N. Kershaw, S. Sidney, K. Yaffe, L. Launer, and A. Zeki Al Hazzouri (2020). “Association of Racial Residential Segregation Throughout Young Adulthood and Cognitive Performance in Middle-aged Participants in the CARDIA Study”. *JAMA Neurology* 77.8, pp. 1000–1007. DOI: [10.1001/jamaneurol.2020.0860](https://doi.org/10.1001/jamaneurol.2020.0860).
- CDC (2024). *Tobacco Product Use Among Adults—United States, 2022: 2022 National Health Interview Survey (NHIS) Highlights*. Tech. rep. Accessed: 2026-01-31. U.S. Department of Health, Human Services, Centers for Disease Control, and Prevention. URL: <https://www.cdc.gov/tobacco/media/pdfs/2024/09/cdc-osh-ncis-data-report-508.pdf>.
- CDC and the Alzheimer’s Association (2007). *The Healthy Brain Initiative: A National Public Health Road Map to Maintaining Cognitive Health*. Chicago, IL: Alzheimer’s Association.
- Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). “Inference on counterfactual distributions”. *Econometrica* 81, pp. 2205–2268.
- Collins, William J. (2021). “The Great Migration of Black Americans from the US South: A guide and interpretation”. *Explorations in Economic History* 80, p. 101382. ISSN: 0014-4983. DOI: <https://doi.org/10.1016/j.eeh.2020.101382>. URL: <https://www.sciencedirect.com/science/article/pii/S001449832030084X>.
- Corral, I., H. Landrine, Y. Hao, L. Zhao, J. L. Mellerson, and D. L. Cooper (2012). “Residential segregation, health behavior and overweight/obesity among a national sample of African American adults”. *Journal of Health Psychology* 17.3, pp. 371–378. DOI: [10.1177/1359105311417191](https://doi.org/10.1177/1359105311417191).
- Crimmins, E. M., J. K. Kim, K. M. Langa, and D. R. Weir (2011). “Assessment of cognition using surveys and neuropsychological assessment: the Health and

- Retirement Study and the Aging, Demographics, and Memory Study". *Journal of Gerontology B: Psychological Sciences and Social Sciences* 66 Suppl 1, pp. i162–i171. DOI: [10.1093/geronb/gbr048](https://doi.org/10.1093/geronb/gbr048).
- Currie, J. (2011). "Inequality at birth: Some causes and consequences". *American economic review* 101.3, pp. 1–22.
- Currie, J. and D. Almond (2011). "Human capital development before age five". In: *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1315–1486.
- DiNardo, J., N.M. Fortin, and T. Lemieux (1996). "Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach". *Econometrica* 64, pp. 1001–1044.
- Do, D. P., R. Frank, and J. Iceland (2017). "Black-white metropolitan segregation and self-rated health: investigating the role of neighborhood poverty". *Social Science & Medicine* 187, pp. 85–92.
- Erreygers, G. and R. Kessels (2013). "Regression-Based Decompositions of Rank-Dependent Indicators of Socioeconomic Inequality of Health". In: *Health and Inequality*. Vol. 21. Research on Economic Inequality, pp. 227–259.
- Erreygers, G. and T. Van Ourti (2011). "Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: A recipe for good practice". *Journal of Health Economics* 30.4, pp. 685–694. DOI: [10.1016/j.jhealeco.2011.04.004](https://doi.org/10.1016/j.jhealeco.2011.04.004).
- Essama-Nssah, B. and P.J. Lambert (2012). "Influence Functions for Policy Impact Analysis". In: *Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber*. Ed. by J.A. Bishop and R. Salas. Vol. 20. Research on Economic Inequality, pp. 135–159.
- Firpo, S., N. M. Fortin, and T. Lemieux (2009). "Unconditional quantile regressions". *Econometrica* 77, pp. 953–973.
- Firpo, S. P., N. M. Fortin, and T. Lemieux (2018). "Decomposing wage distributions using recentered influence function regressions". *Econometrics* 6.2, p. 28.
- Fortin, N.M., T. Lemieux, and S. Firpo (2011). "Decomposition methods in economics". In: *Handbook of Labor Economics*. Ed. by O. Ashenfelter and D. Card. Vol. 4a, pp. 1–102.

- Hampel, F. R. (1968). “Contribution to the theory of robust estimation”. PhD thesis. University of California at Berkeley.
- (1974). “The influence curve and its role in robust estimation”. *Journal of the American Statistical Association* 60, pp. 383–393.
- Heckley, G., U.-G. Gerdtham, and G. Kjellsson (2016). “A general method for decomposing the causes of socioeconomic inequality in health”. *Journal of Health Economics* 48, pp. 89–106.
- Heckman, J., R. Pinto, and P. Savelyev (2013). “Understanding the mechanisms through which an influential early childhood program boosted adult outcomes”. *American Economic Review* 103.6, pp. 2052–2086.
- Hirsch, A. R. (1983). *Making the Second Ghetto: Race and Housing in Chicago, 1940-1960*. Cambridge University Press.
- Hu, Fang and Yan Li (2020). “Overnutrition of children under 5 and women of reproductive age in Egypt”. In: *Scaling Up Nutrition in the Arab Republic of Egypt. Investing in a Healthy Future*. Ed. by Christopher H. Herbst, Amr Elshalakani, Jakub Kakietek, Alaa Hafiz, and Oleg Petrovic. Washington, DC: World Bank.
- IDSC (Egyptian Cabinet Information and Decision Support Center) (2014). *The Cost of Hunger in Egypt: Implications of Child Undernutrition on the Social and Economic Development of Egypt*. Cairo.
- Jarvis, M. J. (Mar. 1996). “The association between having children, family size and smoking cessation in adults”. *Addiction* 91.3, pp. 427–434.
- Johansson, A. K., A. Halling, and The LinQuest Study Group (Sept. 15, 2003). “Does Having Children Affect Adult Smoking Prevalence and Behaviours at Home?” *Tobacco Induced Diseases* 1.1, p. 16. DOI: [10.1186/1617-9625-1-16](https://doi.org/10.1186/1617-9625-1-16).
- Khaled, M., P. Makdissi, and M. Yazbeck (2018). “Income-Related Health Transfers: Principles and Orderings of Joint Distributions of Income and Health”. *Journal of Health Economics* 57, pp. 315–331. DOI: [10.1016/j.jhealeco.2017.08.004](https://doi.org/10.1016/j.jhealeco.2017.08.004).
- Klarman, M. J. (2007). *Unfinished business: Racial equality in American history*. Oxford University Press.
- Kramer, M. R. and C. R. Hogue (2009). “Is Segregation Bad for Your Health?” *Epidemiologic Reviews* 31.1, pp. 178–194. DOI: [10.1093/epirev/mxp001](https://doi.org/10.1093/epirev/mxp001).

- Krieger, N., J. T. Chen, B. Coull, P. D. Waterman, and J. Beckfield (2013). “The unique impact of abolition of Jim Crow laws on reducing inequities in infant death rates and implications for choice of comparison groups in analyzing societal determinants of health”. *American Journal of Public Health* 103.12, pp. 2234–2244.
- Krieger, N., J. L. Jahn, and P. D. Waterman (2017). “Jim Crow and estrogen-receptor-negative breast cancer: US-born black and white non-Hispanic women, 1992-2012”. *Cancer Causes & Control* 28.1, pp. 49–59. DOI: [10.1007/s10552-016-0834-2](https://doi.org/10.1007/s10552-016-0834-2).
- Langa, K. M., D. J. Llewellyn, I. A. Lang, D. R. Weir, R. B. Wallace, M. U. Kabeto, and F. Huppert (2009). “Cognitive health among older adults in the United States and in England”. *BMC Geriatrics* 9, p. 23. DOI: [10.1186/1471-2318-9-23](https://doi.org/10.1186/1471-2318-9-23).
- Langa, K. M., B. L. Plassman, R. B. Wallace, A. R. Herzog, S. G. Heeringa, M. B. Ofstedal, and R. J. Willis (2005). “The aging, demographics, and memory study: Study design and methods”. *Neuroepidemiology* 25, pp. 181–191. DOI: [10.1159/000087448](https://doi.org/10.1159/000087448).
- Liedholm, P., G. Rybo, N.-O. Sjöberg, and L. Sölvell (1975). “Copper IUD — Influence on menstrual blood loss and iron deficiency”. *Contraception* 12.3, pp. 317–325. ISSN: 0010-7824. DOI: [https://doi.org/10.1016/0010-7824\(75\)90091-8](https://doi.org/10.1016/0010-7824(75)90091-8). URL: <https://www.sciencedirect.com/science/article/pii/0010782475900918>.
- Lin, Z., J. Ye, H. Allore, T. M. Gill, and X. Chen (2024). “Early-Life Circumstances and Racial Disparities in Cognition Among Older Adults in the US”. *JAMA Internal Medicine* 184.8, pp. 904–914. DOI: [10.1001/jamainternmed.2024.1132](https://doi.org/10.1001/jamainternmed.2024.1132).
- Machado, J. F. and J. Mata (2005). “Counterfactual decomposition of changes in wage distributions using quantile regression”. *Journal of Applied Econometrics* 20, pp. 445–465.
- Makdissi, P. and M. Yazbeck (2014). “Measuring Socioeconomic Health Inequalities in Presence of Multiple Categorical Information”. *Journal of Health Economics* 34.1, pp. 84–95. DOI: [10.1016/j.jhealeco.2013.11.005](https://doi.org/10.1016/j.jhealeco.2013.11.005).
- Manly, J. J., R. N. Jones, K. M. Langa, L. H. Ryan, D. A. Levine, R. McCammon, S. G. Heeringa, and D. Weir (2022). “Estimating the Prevalence of Dementia and Mild Cognitive Impairment in the US: The 2016 Health and Retirement Study

- Harmonized Cognitive Assessment Protocol Project". *JAMA Neurology* 79.12, pp. 1242–1249. DOI: [10.1001/jamaneurol.2022.3543](https://doi.org/10.1001/jamaneurol.2022.3543).
- Massey, D. S. and N. A. Denton (1988). "The Dimensions of Residential Segregation". *Social Forces* 67.2, 281–315.
- MedlinePlus (2024). *Aging changes in body shape*. Accessed: 2026-02-09. National Library of Medicine (US), National Institutes of Health. URL: <https://medlineplus.gov/ency/article/003998.htm>.
- Meyer, O. L., L. Besser, D. Mitsova, M. Booker, E. Luu, M. Tobias, S. T. Farias, D. Mungas, C. DeCarli, and R. A. Whitmer (2021). "Neighborhood racial/ethnic segregation and cognitive decline in older adults". *Social Science & Medicine* 284, p. 114226. DOI: [10.1016/j.socscimed.2021.114226](https://doi.org/10.1016/j.socscimed.2021.114226).
- Mises, R. von (1947). "On the Asymptotic Distribution of Differentiable Statistical Functions". *Annals of Mathematical Statistics* 18, pp. 309–348.
- MOHP (Ministry of Health and Population) (2012). *Egypt Nutrition Landscape Analysis Report 2012*. Tech. rep. Cairo: Government of Egypt.
- Oaxaca, R. (1973). "Male-female wage differentials in urban labor markets". *International Economic Review* 14, pp. 693–709.
- Ofstedal, M. B., G. G. Fisher, and A. R. Herzog (2005). "Documentation of Cognitive Functioning Measures in the Health and Retirement Study". URL: <http://hrsonline.isr.umich.edu/sitedocs/userg/dr-006.pdf>.
- Piontak, J. R. and M. D. Schulman (2016). "School context matters: the impacts of concentrated poverty and racial segregation on childhood obesity". *Journal of School Health* 86, pp. 864–872.
- Power, M. C., E. E. Bennett, R. W. Turner, N. M. Dowling, A. Ciarleglio, M. M. Glymour, and K. Z. Gianattasio (2021). "Trends in relative incidence and prevalence of dementia across non-Hispanic Black and White individuals in the United States, 2000-2016". *JAMA Neurology* 78.3, pp. 275–284.
- Reddy, K. P. and R. T. Brown (2022). "Association between residential segregation and Black-White disparities in frequent mental distress among older adults". *International Journal of Geriatric Psychiatry* 37.9. DOI: [10.1002/gps.5796](https://doi.org/10.1002/gps.5796).

- Robert, S. A. and E. Ruel (2006). “Racial segregation and health disparities between Black and White older adults”. *Journal of Gerontology B: Psychological Sciences and Social Sciences* 61.4, S203–211. DOI: [10.1093/geronb/61.4.s203](https://doi.org/10.1093/geronb/61.4.s203).
- Rothe, C. (2015). “Decomposing the Composition Effect”. *Journal of Business Economics and Statistics* 33, pp. 323–337.
- Shi, Y., D. Hartley, B. Mazumder, and A. Rajan (2022). “The effects of the Great Migration on urban renewal”. *Journal of Public Economics* 209, p. 104647. ISSN: 0047-2727. DOI: <https://doi.org/10.1016/j.jpubeco.2022.104647>. URL: <https://www.sciencedirect.com/science/article/pii/S0047272722000494>.
- Sokol, J. (2014). *All eyes are upon us: race and politics from Boston to Brooklyn*. Basic Books.
- Sudano, J. J., A. Perzynski, D. W. Wong, N. Colabianchi, and D. Litaker (2013). “Neighborhood racial residential segregation and changes in health or death among older adults”. *Health & Place* 19, pp. 80–88. DOI: [10.1016/j.healthplace.2012.09.015](https://doi.org/10.1016/j.healthplace.2012.09.015).
- Sugrue, T. J. (2008). *Sweet Land of Liberty: The Forgotten Struggle for Civil Rights in the North*. Random House.
- Tang, G. H. and M. Sholzberg (2024). “Iron deficiency anemia among women: An issue of health equity”. *Blood Reviews* 64, p. 101159. ISSN: 0268-960X. DOI: <https://doi.org/10.1016/j.blre.2023.101159>. URL: <https://www.sciencedirect.com/science/article/pii/S0268960X23001297>.
- Vu, H., T. L. Green, and L. E. T. Swan (2024). “Born on the wrong side of the tracks: Exploring the causal effects of segregation on infant health”. *Journal of Health Economics* 95, p. 102876. DOI: [10.1016/j.jhealeco.2024.102876](https://doi.org/10.1016/j.jhealeco.2024.102876).
- Wagstaff, A. (2002). “Inequality Aversion, Health Inequalities and Health Achievement”. *Journal of Health Economics* 21, pp. 627–641.
- Wagstaff, A., P. Paci, and E. van Doorslaer (1991). “On the measurement of inequalities in health”. *Social Science & Medicine* 33, pp. 545–557.
- Walsemann, K. M. and B. A. Bell (2010). “Integrated schools, segregated curriculum: effects of within-school segregation on adolescent health behaviors and educational aspirations”. *American Journal of Public Health* 100.9, pp. 1687–1695. DOI: [10.2105/AJPH.2009.179424](https://doi.org/10.2105/AJPH.2009.179424).

- Walsemann, K. M., N. L. Hair, M. P. Farina, P. Tyagi, H. Jackson, and J. A. Ailshire (2023). “State-level desegregation in the U.S. South and mid-life cognitive function among Black and White adults”. *Social Science & Medicine* 338, p. 116319. DOI: [10.1016/j.socscimed.2023.116319](https://doi.org/10.1016/j.socscimed.2023.116319).
- Weyand, A. C., A. Chaitoff, G. L. Freed, M. Sholzberg, S. W. Choi, and P. T. McGann (June 2023). “Prevalence of Iron Deficiency and Iron-Deficiency Anemia in US Females Aged 12-21 Years, 2003-2020”. *JAMA* 329.24, pp. 2191–2193. ISSN: 0098-7484. DOI: [10.1001/jama.2023.8020](https://doi.org/10.1001/jama.2023.8020). eprint: https://jamanetwork.com/journals/jama/articlepdf/2806540/jama_veyand_2023_ld_230036_1687798729.1443.pdf. URL: <https://doi.org/10.1001/jama.2023.8020>.
- Williams, A. D., M. Wallace, C. Nobles, and P. Mendola (2018). “Racial residential segregation and racial disparities in stillbirth in the United States”. *Health & Place* 51, pp. 208–216.
- Williams, D. R. and C. Collins (2001). “Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health”. *Public Health Reports* 116.5, pp. 404–416. DOI: [10.1093/phr/116.5.404](https://doi.org/10.1093/phr/116.5.404).
- Williams, D. R., J. A. Lawrence, and B. A. Davis (2019). “Racism and health: evidence and needed research”. *Annual Review of Public Health* 40.1, pp. 105–125.
- Yaffe, K., C. Falvey, T. B. Harris, A. Newman, S. Satterfield, A. Koster, H. Ayonayon, and E. Simonsick (2013). “Effect of socioeconomic disparities on incidence of dementia among biracial older adults: prospective study”. *BMJ* 347, f7051. DOI: [10.1136/bmj.f7051](https://doi.org/10.1136/bmj.f7051).