CAHIER DE RECHERCHE #2403E
Département de science économique
Faculté des sciences sociales
Université d'Ottawa

WORKING PAPER #2403E
Department of Economics
Faculty of Social Sciences
University of Ottawa

A modeling approach to decomposing changes in health concentration curves[*]

Khadija Bchi[†], Paul Makdissi[‡], and Myra Yazbeck[§]

May 2024

[*] We thank Louis-Philippe Béland, Myra Mohnen, Louis-Philippe Morin, and two anonymous referees for their comments on previous versions of this work.
[†] Department of Economics, University of Ottawa, Canada; Email: kbchi030@uottawa.ca
[‡] Department of Economics, University of Ottawa, Canada; Email: paul.makdissi@uottawa.ca
[§] Department of Economics, University of Ottawa, Canada; Email: myazbeck@uottawa.ca

Abstract

This paper proposes a decomposition approach for health concentration curves. Decomposing changes in health concentration curves gives additional insight compared to decomposing a single index such as the health concentration index. First, the results would be valid for a comprehensive set of indices. Second, and more importantly, it allows for identifying heterogeneous effects along socioeconomic ranks. We use inverse propensity weighting for the overall decomposition. We use multiple recentered influence function regressions on a grid of points to identify the impact of specific covariates. We weight these regressions by the inverse propensity score of the observations to correct for errors due to departure from linearity. The paper also derives the expressions of the recentered influence functions of the relative and absolute health concentration curves since the literature does not offer the expression of these recentered influence functions. We offer an empirical illustration using information on cigarette consumption from the National Health Interview Survey of 2000 and 2020.

Key words: *Counterfactual, inverse probability weighting, recentered influence function, health concentration curves*

JEL Classification: I10, D63

1 Introduction

Measuring socioeconomic health inequality and health achievement (shortfall) is an essential exercise from a public health perspective, as it enables us to assess critical aspects such as access to basic healthcare services, maternal and child health outcomes, and variations in life expectancy among different socioeconomic groups. One prevalent approach, is the index-based approach which consists in choosing a specific parametric structure for the inequality measure (see Wagstaff, Paci, and van Doorslaer, 1991; Wagstaff, 2002; Erreygers and Van Ourti, 2011; van Doorslaer and Van Ourti, 2011). An index-based approach can produce a complete ordering of the joint distributions of health and income. However, the result for some of the comparisons may hinge on the selected mathematical structure of the index. Particularly, the ranking may change when the analyst picks another mathematical structure for the index.

Another approach, is the dominance approach. It aims at identifying rankings of distributions that would be robust to all indices belonging to a broad class of indices (see Makdissi and Yazbeck, 2014; Khaled, Makdissi, and Yazbeck, 2018 and 2023). This involves comparing two distributions' health concentration curves. For each social rank, the relative health concentration curve[1] illustrates the cumulative share of the total health of all individuals below and up to this social rank. If these two curves do not intersect, the distribution corresponding to the higher curve is considered to have less health inequality across all indices in the pre-specified class of indices. A similar condition for health achievement (shortfall) indices holds with non-intersection of the absolute health concentration curves (relative health concentration curve times average health status).[2]

[1]The relative health concentration curve is often referred to as the health concentration curve. This paper uses relative to distinguish from the absolute health concentration curve.

[2]For historical reasons in the literature of income inequality, the absolute concentration curve is often referred to as the generalized concentration curve, including in Makdissi and Yazbeck (2014) and in Khaled, Makdissi, and Yazbeck (2018, 2023). In other works, such as Schechtman, Shelef, Yitzhaki and Zitikis (2008), the same curve has been referred to as the absolute concentration curve. Since we believe that using

O'Donnell and Van Ourti (2021) recently suggested that analysts might utilize these dominance approaches to evaluate health equity impacts and trade-offs when performing a distributional cost-effectiveness analysis. This recommendation is pertinent to all forms of health policy assessments. Nevertheless, when estimating the treatment effect of a public health policy or a new medical treatment, the analyst often faces the task of contrasting a population's existing relative or absolute health concentration curve against a hypothetical scenario (counterfactual) – one without the proposed health policy or medical treatment. Unfortunately, by the construction of the potential outcome model, information at the individual observation level is only available for one of the two options: treated or non-treated. In this sense, this issue is often akin to as a missing data problem. This type of missing data problem is not uncommon in various analyses. For instance, when an analyst aims to construct a disaggregated profile of health inequality at a local level, they often rely on a model derived from a broader national survey (with information on health, income, and some demographic characteristics) to project local-level curves. However, this is complicated by national censuses typically lacking specific health and income data, even though they might provide other demographic specifics, a challenge discussed in studies by Elbers, Lanjouw, and Lanjouw (2003 and 2005). Another scenario is evident in the well-known Oaxaca-Blinder decomposition analyses. In all these situations, employing an econometric model capable of predicting a hypothetical scenario's relative or absolute health concentration curve is imperative.

Heckley, Gerdtham, and Kjellsson (2016) and Kessel and Erreygers (2019) offer approaches for modeling health concentration indices. Abu Ismail, Gantner, Makdissi, and Yazbeck (2020) have further modified the technique initiated by Heckley, Gerdtham, and Kjellsson (2016) for application to the health achievement (shortfall) index. Nonetheless,

"absolute" better depicts the nature of the curve, we chose to use this term in this paper.

to our knowledge, the literature does not yet offer a framework for modeling the relative and absolute health concentration curve. Proposing a modeling approach to changes in the health concentration curve is important for two reasons. First, analysts use health concentration curves to identify dominance results, i.e., results that remain valid for a wide set of indices. Second, the health concentration curves also allow for identifying potential heterogeneous impacts along socioeconomic ranks.

This paper aims to fill this gap by drawing from insights from Firpo and Pinto (2016) and Heckley, Gerdtham, and Kjellsson (2016). We rely on Firpo and Pinto's (2016) inverse propensity weight approach for the overall decomposition of changes in the health concentration curve. In order to identify the impact at the covariate level, we follow Heckley, Gerdtham, and Kjellsson (2016) and use weighted RIF-regressions. We extend their approach by running multiple regressions on a grid of points and using inverse propensity reweighting to estimate our regression models to account for the potential error in the linear approximation inherent to RIF-regressions. The motivation for using inverse propensity weights in estimating the regression parameters lies in the fact that Firpo, Fortin, and Lemieux (2009) have shown that the parameters of a RIF-regression identify the impact of a marginal change in the distribution of a covariate. However, often, counterfactuals involve non-marginal changes in the distribution of covariates. In this case, Rothe (2015) has shown that in the case of non-linear functionals, RIF-regression yields a measurement error.

The remainder of this paper runs as follows. First, Section 2 presents some essential background from the literature on socioeconomic health inequality. Then, Section 3 presents the measurement framework of the dominance approach to socioeconomic health inequality comparisons. This section also proposes a decomposition approach and derives the expressions of the recentered influence functions for the relative and absolute health concentration curves. Next, in Section 4, we provide an empirical demonstration of the modeling approach by decomposing the change in the relative and absolute health concentration curves of cigarette consumption between 2000 and 2020 in the US. Finally, in Section 5, we conclude.

2 Measuring socioeconomic health inequality

Assume we have two random variables: health, H, and income, Y. Assume that these two variables are absolutely continuous with a joint distribution F_{HY} defined over the support $[0, h_{\max}] \times \Re_{+}$. Let us use the usual notation and denote by $F_{H|Y}$ the cumulative distribution of health conditional on income, by F_H and F_Y, the unconditional cumulative distributions of health and income, and by $F_Y^{-1}(q)$, the q-th quantile of the distribution of income.

Since the seminal paper of Wagstaff, Paci, and van Doorslaer (1991), researchers in health inequality have been using the relative health concentration curve to offer a graphical representation of the distribution of expected health by socioeconomic status. For each social rank $q \in [0,1]$, the relative health concentration curve displays the cumulative share of total health of the q poorest individuals. Figure 1 displays the relative health concentration curves associated with two joint distributions of health and income, $F_{H_0Y_0}$ and $F_{H_1Y_1}$. The curve $C_R(q; F_{H_0Y_0})$ (in red) depicts a situation in which the health outcome is more concentrated at higher income quantiles and the curve $C_R(q; F_{H_1Y_1})$ (in green), a situation in which the health outcome is more concentrated at lower income quantiles. Wagstaff, Paci, and van Doorslaer (1991) suggested using the canonical health concentration index $CI(F_{HY})$ as a measure of socioeconomic health inequality. This index, which takes values between -1 and 1, equals twice the surface between the line of perfect socioeconomic health equality (the 45-degree line) and the relative health concentration curve. When the index has a positive value, as in $CI(F_{H_0Y_0})$, socioeconomic health inequalities are pro-rich. When it takes a negative value, as in $CI(F_{H_1Y_1})$, it signals that these socioeconomic inequalities are pro-poor. It is also possible to apply these tools to ill-health variables. In this case, one should reverse the interpretation of the pro-poor/pro-rich cases.

Wagstaff (2002) points to two issues associated with using the canonical health concentration index. First, it imposes a specific level of aversion to socioeconomic health inequality. Second, it ignores differences in the level of average health outcomes. For this reason, he argues that a health achievement (shortfall) index should be considered as well for policy recommendations. In combining both the average health outcome with its relative socioeconomic inequality, the information provided by such indices would better inform decision-makers' social preferences concerning health policy.

Makdissi and Yazbeck (2014) and Khaled, Makdissi, and Yazbeck (2018, 2023) build on Wagstaff's argument that one should aim to consider potential differences in socioeconomic aversion. Instead of relying on a specific index or a specific class of indices, they propose the use of positional dominance conditions and tests to identify rankings of health achievement (shortfall) and socioeconomic health inequality that would remain the same for all rank-dependent health achievement (shortfall) indices or to any rank-dependent socioeconomic health inequality indices. In such cases, the result is independent of the specific mathematical structure imposed when using a specific class of indices.

In this paper, we focus only on the two most important results of these papers. The first is that when the relative health concentration curves of two distributions do not intersect, the relative health concentration curve above the other is associated with a lower level of socioeconomic health inequality. This result remains the same for any rank-dependent socioeconomic health inequality index in line with Bleichrodt and van Doorslaer's (2006) *principle of income-related health transfer*. This normative principle implies that a "hypothetical" transfer of health from one person to another at a lower socioeconomic rank is considered a social improvement.

The second result relies on comparisons of absolute health concentration curves. The absolute health concentration curve of a distribution is equal to the relative health concentration curve multiplied by the average health status, $C_A(q; F_{HY}) = \mu_H C_R(q; F_{HY})$. When the absolute health concentration curves of two distributions do not intersect, the absolute health concentration curve above the other is associated with a higher level of health achievement (shortfall) than the other. This result is valid for any rank-dependent health achievement (shortfall) index in line with Bleichrodt and van Doorslaer's (2006) *principle of income-related health transfer*.[3]

3 Estimating and modeling health concentration curves

3.1 Estimating health concentration curves

Assume that an analyst is interested in health achievement or shortfall. Also, assume that this analyst aims at comparing health achievement or shortfall using absolute health concentration curve rankings. The formal mathematical definition of an absolute health concentration curve associated with F_{HY} is

$$C_A(q; F_{HY}) = \int_0^{F_Y^{-1}(q)} \int_0^{h_{\max}} h f_{HY}(h, y) \mathrm{d}h \mathrm{d}y. \tag{1}$$

Utilizing a rank-dependent health achievement (shortfall) approach proves to be advantageous for health policy recommendation because it combines the average health outcome and the socioeconomic inequalities in its distribution. However, to gain a more comprehensive understanding, it is also essential to conduct analyses specifically focusing on the inequalities present in health outcomes across different socioeconomic levels. Comparing relative health concentration curves allows for identifying relative socioeconomic health inequality rankings. A relative health concentration curve is defined mathematically as follows:

[3] These two cases correspond to the "second-order" curves in Makdissi and Yazbeck (2014) and Khaled, Makdissi, and Yazbeck (2018, 2023).

$$C_R(q; F_{HY}) = \frac{C_A(q; F_{HY})}{\mu_H}. \tag{2}$$

3.2 Modeling health concentration curves

In applied health economics research, the analysts find themselves needing to compare the estimated effects on an outcomes of interest with estimated effects using an alternative counterfactual scenario. One prominent example of such as scenario is the Neyman–Rubin potential outcome framework (Neyman, 1924; Rubin, 1974) developed for a univariate outcome variable. In this paper, we extend this framework to a bivariate outcome variable. Our focus in this paper is on the joint health/income distribution of two groups, group 1 and group 0. Subsequent to this, our discussion operates under the assumption that we have a random assignment procedure of the units of observation between these two groups. In many applied contexts, the analyst must ensure that the identification assumptions are such that their setting mimics this type of random assignment.

For an individual i, let (H_{1i}, Y_{1i}) be the joint health and income outcome if the individual is assigned to group 1 and let (H_{0i}, Y_{0i}) be the joint health and income outcome if the individual is assigned to group 0. Since the same individual i can either be observed in group 1 or in group 0 but not in both groups at the same time, we never have data on the same individual in both groups. Therefore, for each individual the potential outcome framework is defined as follows: $(H_i, Y_i) = (H_{1i} \cdot T_i, Y_{1i} \cdot T_i) + (H_{0i} \cdot (1 - T_i), Y_{0i} \cdot (1 - T_i))$, where $T_i = 1$ if individual i is observed in group 1, and $T_i = 0$ if individual i is observed in group 0. When individuals are assigned, we observe one of the two potential joint outcomes (H_{1i}, Y_{1i}) or (H_{0i}, Y_{0i}).

The health policy analyst is often interested in evaluating the impact of the assignment to particular group, for instance group 1, on an outcome of interest, for instance, health inequality. In such cases, using health concentration curve dominance requires estimating a counterfactual health concentration curve for individuals in group 1 had they been assigned to group 0. In the context of this paper's counterfactual analysis we assume that we have two groups labeled by $T \in \{0,1\}$. Each group, T, has a vector X_T with support $\mathscr{X}_T$ of covariates and the joint potential outcome variable, (H,Y). Let us denote by $F_{H_{1T}Y_{1T}X_T}$ the joint cumulative distribution of potential health and income, and covariates of group $T \in \{0,1\}$ if assigned to $T = 1$ and by $F_{H_{0T}Y_{0T}X_T}$ the joint cumulative distribution of potential health and income, and covariates of group $T \in \{0,1\}$ if assigned to $T = 0$.

We only observe outcomes for group $T = 1$ assigned to group 1 and group $T = 0$ assigned to group 0. For each of these groups, we have a direct observation of the absolute health concentration curves:

$$C_A(q; F_{H_T Y_T X_T}) = \int_{\mathscr{X}} \int_0^{F_{Y_T}^{-1}(q)} \int_0^{h_{\max}} h f_{H_T Y_T|X}(h, y|X = x) \mathrm{d}h \mathrm{d}y \mathrm{d}F_{X_T}(x), \quad T \in \{0,1\}. \tag{3}$$

For policy evaluation purposes, one often needs information on the counterfactual absolute health concentration curve of population $T = 1$; in the hypothetical case where it would have been assigned to group 0. Let us denote this counterfactual absolute health concentration curve by $C_{A\langle 0|1\rangle}(q)$. This counterfactual curve is never observed. Identifying this counterfactual curve using the information of both groups $T = 0$ and 1 requires the following assumptions.

Assumption 1. ***Ignorability:*** *Let (T, X, ε) have a joint distribution. $\varepsilon \perp\!\!\!\perp T|X = x$, for all $x \in \mathscr{X}$.*

Assumption 2. ***Overlapping Support****: For all $x \in \mathscr{X}$, $p(x) = Pr[T = 1|X = x] < 1$ and $Pr[T = 1] > 0$.*

Under Assumptions 1 and 2, the counterfactual absolute health concentration curve is identifiable as:

$$C_{A\langle 0|1\rangle}(q) = \int_{\mathscr{X}} \int_0^{F_{Y_{\langle 0|1\rangle}}^{-1}(q)} \int_0^{h_{\max}} h f_{H_0 Y_0|X}(h, y|X = x) \mathrm{d}h \mathrm{d}y \mathrm{d}F_{X_1}(x), \tag{4}$$

where $F^{-1}_{Y_{\langle 0|1\rangle}}(q) = \inf\left\{y|F_{Y_{\langle 0|1\rangle}}(y) \geq q\right\}$ and

$$F_{Y_{\langle 0|1\rangle}}(y) = \int_{\mathscr{X}} \int_0^y h f_{Y_0|X}(s|X = x)\mathrm{d}s\mathrm{d}F_{X_1}(x). \tag{5}$$

Similarly, one may also want to build a counterfactual relative health concentration curve:

$$C_{R\langle 0|1\rangle}(q) = \frac{C_{A\langle 0|1\rangle}(q)}{\mu_{H\langle 0|1\rangle}}, \tag{6}$$

where

$$\mu_{H\langle 0|1\rangle} = \int_{\mathscr{X}} \int_0^{h_{\max}} h f_{H_0|X}(h|x)\mathrm{d}h\mathrm{d}F_{X_1}(x). \tag{7}$$

In what follows, we will use the notation $\langle 1|1\rangle$ and $\langle 0|0\rangle$ when referring to functionals of the observed distributions $F_{H_1Y_1X_1}$ and $F_{H_0Y_0X_0}$.

3.3 Estimating counterfactual health concentration curves: a reweighting approach

The literature in applied econometrics offers some approaches to model a functional of a counterfactual distribution if one is interested in the overall impact of the assignment to group $T = 1$. Part of the literature proposes to model the distribution and then use the functional of the modeled distribution (e.g., Chernozhokov, Fernandez-Vál, and Melly (2013). Alternatively, one can follow DiNardo, Fortin, and Lemieux (1996) and use inverse probability weighting to model the counterfactual distribution. This last methodological approach is the avenue chosen by Firpo and Pinto (2016) and Firpo, Fortin, and Lemieux (2018). In this paper we use Firpo, Fortin, and Lemieux (2018) to model the counterfactual distribution. Specifically, under Assumptions 1 and 2, one can use their reweighting approach to estimate the counterfactuals $C_{A\langle 0|1\rangle}(q)$ and $C_{R\langle 0|1\rangle}(q)$.

Assume that we have a sample of size N, with N_1 observations of population with $T = 1$ and N_0 observations of population with $T = 0$. Let p be the proportion of observations belonging to group $T = 1$ in the joint sample of $T = 0$ and 1. In the first step, we must estimate a propsensity score using a binary model of $p(x) = \Pr[T = 1|X = x]$. The predicted values of this model, can then be used to estimate the following three weight functions proposed by Firpo and Pinto (2016):

$$\widehat{\omega}_{\langle 1|1 \rangle}(T_i, x_i) = \frac{T_i}{\widehat{p}}, \tag{8}$$

$$\widehat{\omega}_{\langle 0|0 \rangle}(T_i, x_i) = \frac{1 - T_i}{1 - \widehat{p}}, \tag{9}$$

and

$$\widehat{\omega}_{\langle 0|1 \rangle}(T_i, x_i) = \left(\frac{\widehat{p}(x_i)}{1 - \widehat{p}(x_i)} \right) \cdot \left(\frac{1 - T_i}{\widehat{p}} \right). \tag{10}$$

We follow Firpo, Fortin, and Lemieux (2018) and normalize these weights as

$$\widehat{\omega}^*_{\langle 1|1 \rangle}(T_i, x_i) = \frac{\widehat{\omega}_{\langle 1|1 \rangle}(T_i, x_i)}{\sum_{i=1}^{N} \widehat{\omega}_{\langle 1|1 \rangle}(T_i, x_i)}, \tag{11}$$

$$\widehat{\omega}^*_{\langle 0|0 \rangle}(T_i, x_i) = \frac{\widehat{\omega}^*_{\langle 0|0 \rangle}(T_i, x_i)}{\sum_{i=1}^{N} \widehat{\omega}^*_{\langle 0|0 \rangle}(T_i, x_i)}, \tag{12}$$

and

$$\widehat{\omega}^*_{\langle 0|1 \rangle}(T_i, x_i) = \frac{\widehat{\omega}_{\langle 0|1 \rangle}(T_i, x_i)}{\sum_{i=1}^{N} \widehat{\omega}_{\langle 0|1 \rangle}(T_i, x_i)}. \tag{13}$$

One can then use these weight functions similarly to survey weights to estimate the desired relative and absolute health concentration curves and their counterfactuals.

The steps described above suggest that estimating the counterfactual health concentration curves is akin to estimating these curves from the observed distribution once the binary probability model is modeled. It is important to note that since the weight functions are estimated values, we need to reestimate these weights at each iteration of a bootstrap procedure.

It is often interesting to assess the impact of being assigned to $T = 1$ on health achievement (shortfall) and socioeconomic health inequality. Within this framework, it is often interesting to decompose the change in the overall difference in health achievement (shortfall) and socioeconomic health inequality into two components, following the Oaxaca-Blinder's decomposition approach. The first component, termed the structural effect, quantifies the portion of the overall difference attributed to the assignment to $T = 1$. For example, if the assignment to $T = 1$ corresponds to receiving a non-placebo medical treatment or undergoing policy intervention, this component can be interpreted as a treatment effect. Conversely, when being assigned to $T = 1$ signifies membership in a specific gender or racial category, it is commonly interpreted by analysts as the manifestation of the impact of discrimination. The second element of the decomposition, known the composition effect, isolates the portion of the overall difference attributable to the difference in the distribution of covariates. Let $\widehat{\Delta C}_A(q) = \widehat{C}_{A\langle 1|1\rangle}(q) - \widehat{C}_{A\langle 0|0\rangle}(q)$ and $\widehat{\Delta C}_R(q) = \widehat{C}_{R\langle 1|1\rangle}(q) - \widehat{C}_{R\langle 0|0\rangle}(q)$. Using the above methodology allows us to perform these decompositions:

$$\widehat{\Delta C}_A(q) = \underbrace{\left[\widehat{C}_{A\langle 1|1\rangle}(q) - \widehat{C}_{A\langle 0|1\rangle}(q)\right]}_{\text{structural effect}} + \underbrace{\left[\widehat{C}_{A\langle 0|1\rangle}(q) - \widehat{C}_{A\langle 0|0\rangle}(q)\right]}_{\text{composition effect}} \tag{14}$$

$$\widehat{\Delta C}_R(q) = \underbrace{\left[\widehat{C}_{R\langle 1|1\rangle}(q) - \widehat{C}_{R\langle 0|1\rangle}(q)\right]}_{\text{structural effect}} + \underbrace{\left[\widehat{C}_{R\langle 0|1\rangle}(q) - \widehat{C}_{R\langle 0|0\rangle}(q)\right]}_{\text{composition effect}} \tag{15}$$

Since the decomposition offered in equations (14) and (15) pertains to health concentration curves, the results can be interpreted in alignment with the principles of dominance conditions. For instance, if $\widehat{C}_{A\langle 1|1\rangle}(q) - \widehat{C}_{A\langle 0|1\rangle}(q) \geq 0$ for all $q \in [0, 1]$, then the structural effect would amplify health achievement (shortfall) across any rank-dependent health achievement (shortfall) indices in line with the *principle of income-related health transfer* posited by Bleichrodt and van Doorslaer's (2006). Analogously,, should $\widehat{C}_{R\langle 1|1\rangle}(q) - \widehat{C}_{R\langle 0|1\rangle}(q) \geq 0$ for all $q \in [0, 1]$, the structural effect mitigate socioeconomic health inequality for any rank-dependent socioeconomic health inequality index adhering to Bleichrodt and van Doorslaer's (2006) *principle of income-related health transfer*. Corresponding interpretations are applicable for the composition effect.

3.4 Estimating counterfactual health concentration curves: a recentered influence function regression approach

In instances where the examination is centered around the specific structural influence of each individual covariate in X on the impact of assignment to group $G = 1$, it is appropriate to employ an alternative approach that directly models the functional of the distribution. A methodological solution for such a direct approach has been provided by Firpo, Fortin, and Lemieux (2009) in the form of the recentered influence function (RIF) regression approach, that aims to directly model the functional of a univariate distribution. Subsequently, the extension of this methodology to the bivariate context, particularly the health concentration index, was executed by Heckley, Gerdtham, and Kjellsson (2016). Given that both the absolute and relative health concentration curves also represent functionals of a bivariate distribution of health and income, this section aims to incorporate the insights provided by Heckley, Gerdtham, and Kjellsson (2016) to the context of concentration curves. The intention is to extend the approach of Firpo, Fortin, and Lemieux (2009) for estimating a model of $C_A(q; F_{HY})$ and $C_R(q; F_{HY})$.

To illustrate the methodology, let us examine the instance of the absolute health concentration curve. The influence function of the absolute health concentration curve of observation i, $IF(h_i, y_i; C_A(q; F_{HY}), F_{HY})$, represents the effect on $C_A(q; F_{HY})$ attributable to an infinitesimal contamination of F_{HY} at point mass (h_i, y_i). To formally derive this influence function, the Dirac distribution function is considered, characterized by a degenerate probability mass at (h_i, y_i):

$$\delta_{HY}(h, y; h_i, y_i) = \begin{cases} 0 & \text{if } h < h_i \text{ or } y < y_i \\ 1 & \text{if } h \geq h_i \text{ and } y \geq y_i \end{cases} . \tag{16}$$

Also, consider $\widetilde{F}_{HY}$, a mixture of distribution F_{HY} and δ_{HY}, $\widetilde{F}_{HY} = (1-t)F_{HY} + t\delta_{HY}$. The influence function of the qth coordinate of the absolute health concentration curve is given by

$$
\begin{aligned}
IF(h_i, y_i; C_A(q; F_{HY}), F_{HY}) &= \lim_{t \to 0} \frac{C_A\left(q; \widetilde{F}_{HY}\right) - C_A(q; F_{HY})}{t} \\
&= \frac{\partial}{\partial t} C_A(q; [1-t]F_{HY} + t\delta_{HY}(h, y; h_i, y_i))\Big|_{t=0}. \quad (17)
\end{aligned}
$$

This influence function is a Gâteaux derivative of the absolute health concentration curve's qth coordinate. It quantifies the marginal effect on $C_A(q; F_{HY})$ arising from a small perturbation of the joint distribution function F_{HY} at (h_i, y_i).

Firpo, Fortin, and Lemieux (2009) utilize the property that, by definition, the expected value of an influence function is equal to zero. Consequently, they suggest incorporating the functional value to the influence function to generate a recentered influence function. In the context of our study, this implies the addition of the value of the qth coordinate of the absolute health concentration curve to its corresponding influence function:

$$
RIF(h_i, y_i; C_A(q; F_{HY}), F_{HY}) = C_A(q; F_{HY}) + IF(h_i, y_i; C_A(q; F_{HY}), F_{HY}). \quad (18)
$$

The $\mathrm{E}[IF(h, y; C_A(q; F_{HY}), F_{HY})]$ is zero, which indicates that the expected value of this RIF corresponds to the qth point on the absolute health concentration curve. Additionally, the influence function is also the Gâteaux derivative of $C_A(q; F_{HY})$. Therefore, $\mathrm{E}[RIF(h, y; C_A(q; F_{HY}), F_{HY})]$ will encompass the primary two terms of a von Mises (1947) approximation of $C_A(q; F_{HY})$. This means the impact of a change in some exogenous variable on the RIF is a first-order approximation of the actual impact on the q-th coordinate of the absolute health concentration curve. In this context, Firpo, Fortin, and Lemieux (2009) show that the coefficient of a regression gives the impact on the functional of interest arising from a marginal change in the distribution of covariates, F_X.

In modeling a counterfactual functional, it is typical to consider non-marginal changes in the distribution of covariates, denoted as F_X. Within this framework, applying a RIF-regression approach involves constructing a functional which could be non-linear, such as the health concentration curves, by utilizing a linear model of iterated expectations of RIFs. Rothe (2015) argues that it is essential to recognize deviations from linearity in such instances and the necessity to quantify the error arising from such deviations. In this context, Firpo, Fortin, and Lemieux (2018) recommend using the weights functions from the prior section into the regression model to estimate the error stemming from the violation of the linearity assumption.

In this paper, we adopt the methodology proposed by Firpo, Fortin, and Lemieux (2018). The implementation of the estimation is structured into three stages. The first stage consists in estimating the binary model and the weights $\widehat{\omega}^*_{\langle 1|1\rangle}(T_i, x_i)$, $\widehat{\omega}^*_{\langle 0|0\rangle}(T_i, x_i)$, and $\widehat{\omega}^*_{\langle 0|1\rangle}(T_i, x_i)$ as in the preceding section.

Subsequently, in the second stage, we estimate the values of the recentered influence function for each observation. Given the absence of the literature on the expressions of the recentered influence functions for the health concentration curves, we resort to the approach of Heckley, Gerdtham, and Kjellsson (2016) to develop these new expressions (comprehensive details 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*

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

Its recentered influence function is given by

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

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

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

Its recentered influence function is given by

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

Examination of Proposition 1 and Corollary 1 reveals that to compute the RIF of both the relative and absolute health concentration curves, one must estimate a non-parametric regression model of $\mathrm{E}[\,H|\,Y]$. Given that selecting the bandwidth for a non-parametric estimation aims at minimizing the mean integrated squared error (MISE), there will be inherent bias in these estimates.[4] In the empirical section of this paper, we will visually compare the health concentration curves obtained by averaging the RIF of the observations with those that are directly estimated. For both absolute and relative health concentration curves, three vectors of recentered influence functions must be estimated. There vectors are estimated by applying the weights $\widehat{\omega}^*_{\langle 1|1\rangle}(T_i, x_i)$, $\widehat{\omega}^*_{\langle 0|0\rangle}(T_i, x_i)$, and $\widehat{\omega}^*_{\langle 0|1\rangle}(T_i, x_i)$ on the observations.

In the third stage, the task is to estimate weighted OLS, $\mathrm{E}\left[RIF|X\right]$, for each of the three vectors of RIF computed in the preceding stage. Here $\widehat{\beta}^{C_A(q)}_{\langle t'|t\rangle}$ the vector of parameters estimated for the model of the absolute health concentration curves, using the weights $\widehat{\omega}^*_{\langle t'|t\rangle}(T_i, x_i)$, $(t', t) \in \{(1,1),(0,0),(0,1)\}$. The parameters $\widehat{\beta}^{C_R(q)}_{\langle t'|t\rangle}$ are defined in a corresponding manner. Given the health concentration curves are defined across the entire range of $q \in [0, 1]$, one must estimate a series of OLS models on a grid of selected points within this range. Note also that since both the weights and the RIFs values are estimated values,

[4] Firpo, Fortin, and Lemieux (2009) face the same issue for the estimation of the RIF of the unconditional quantile function, which requires a kernel density estimation of the density of income.

they necessitate re-estimation during each bootstrap iteration to procure an estimate of the standard errors.

We can use the model to perform a detailed decomposition. Let $\overline{X}_{\langle t'|t\rangle}$ be the vectors of weighted averages of covariates using the weights $\widehat{\omega}^{*}_{\langle t'|t\rangle}(T_i, x_i)$. The overall difference between population 1 and population 0 absolute health concentration curve can be estimated using

$$\begin{aligned}\widehat{\Delta C}_A(q) &= \underbrace{\overline{X}'_{\langle 1|1\rangle}\cdot\left(\widehat{\beta}^{C_A(q)}_{\langle 1|1\rangle}-\widehat{\beta}^{C_A(q)}_{\langle 0|1\rangle}\right)}_{\text{Structural effect}} + \underbrace{\left(\overline{X}_{\langle 1|1\rangle}-\overline{X}_{\langle 0|1\rangle}\right)'\cdot\widehat{\beta}^{C_A(q)}_{\langle 0|1\rangle}}_{\text{Misspecification of the reweighting model}} \\ &\quad + \underbrace{\left(\overline{X}_{\langle 0|1\rangle}-\overline{X}_{\langle 0|0\rangle}\right)'\cdot\widehat{\beta}^{C_A(q)}_{\langle 0|0\rangle}}_{\text{Composition effect}} + \underbrace{\overline{X}'_{\langle 0|1\rangle}\cdot\left(\widehat{\beta}^{C_A(q)}_{\langle 0|1\rangle}-\widehat{\beta}^{C_A(q)}_{\langle 0|0\rangle}\right)}_{\text{Departure from linearity}}, \qquad (19)\end{aligned}$$

$\widehat{\Delta C}_R(q)$ can be decomposed analogously.

Equation (19) allows us to identify both the structural and composition effect of each covariate $\ell \in \{1, \ldots, L\}$. The structural effect of variable ℓ is expressed as $\overline{X}_{\ell\langle 1|1\rangle}\cdot\left(\widehat{\beta}^{C_A(q)}_{\ell\langle 1|1\rangle}-\widehat{\beta}^{C_A(q)}_{\ell\langle 0|1\rangle}\right)$. The composition effect, on the other hand, is represented by $\widehat{\beta}^{C_A(q)}_{\ell\langle 0|0\rangle}\cdot\left(\overline{X}_{\ell\langle 0|1\rangle}-\overline{X}_{\ell\langle 0|0\rangle}\right)$. While this flexibility offers more insights than the method outlined in the previous section, it also entails additional complexities. It requires the estimation of two additional elements. The first error component in equation (19) is linked to the misspecification of the reweighting model and influences the decomposition's structural term. If the reweightng model is correctly specified, $\overline{X}_{\langle 1|1\rangle}$ should be very close to $\overline{X}_{\langle 0|1\rangle}$. This error term affects both the reweighting and RIF decomposition approaches. Nonetheless, this error term is expected to decrease as the size of the dataset increases. As will be demonstrated in the empirical application, even with a simple parametric logit model, this error component is generally insignificant, and its relative magnitude is minimal when does manifest significance.

The second error term is linked with deviations from linearity and impacts the composition term of the decomposition. If the functionals under consideration were linear, then the estimated parameters $\widehat{\beta}^{C_A(q)}_{\langle 0|1\rangle}$ and $\widehat{\beta}^{C_A(q)}_{\langle 0|0\rangle}$ would be close enough as both models are estimated under assignment to group 0. Given the non linear nature of the health concentration curves in our framework, this error term may be non marginal. This error term does not impact the reweighting approach presented in the preceding section. However, the reweighting approach does not allow for identifying each covariate's composition and structural effect.

Similarly to the overall decomposition in section 3.3, dominance interpretations can be applied to analyze the effect of each covariate. For instance, if for all $q \in [0,1]$, $\left(\widehat{\beta}^{C_A(q)}_{\ell\langle 1|1\rangle} - \widehat{\beta}^{C_A(q)}_{\ell\langle 0|1\rangle}\right) \cdot \overline{X}_{\ell\langle 1|1\rangle} \geq 0$ holds true, then the change in the "return" to covariate ℓ would lead to an increased health shortfall for all rank-dependent health shortfall indices in line with the principles of *principle of income-related health transfer* by Bleichrodt and van Doorslaer's (2006). Similarly, if $\left(\widehat{\beta}^{C_R(q)}_{\ell\langle 1|1\rangle} - \widehat{\beta}^{C_R(q)}_{\ell\langle 0|1\rangle}\right) \cdot \overline{X}_{\ell\langle 1|1\rangle} \geq 0$ for all $q \in [0,1]$, then the structural effect would lead to a decreased socioeconomic health inequality for all rank-dependent socioeconomic health inequality indices in line with Bleichrodt and van Doorslaer's (2006) *principle of income-related health transfer*. Similar interpretations hold for the composition effect.

4 Empirical analysis: Decomposing changes in health shortfall and socioeconomic health inequality in cigarette consumption in the USA between 2000 and 2020

In this study we conduct an empirical demonstration using data from National Health Interview Survey for the years 2000 and 2020 to showcase the practical utility of the methodology introduced in this paper. This empirical demonstration, focuses on the decomposition of the change in both the absolute and relative health concentration curves across these two decades with a focus on an important health metric often examined in health economics: daily cigarette consumption. Our decision to analyze cigarette consumption for 2000 and

2020 is driven by two primary considerations. To begin with, the span between these two years offers a unique opportunity to observe significant shifts in the distribution of the studied variables. Moreover, there was a notable evolution in smoking habits during this interval. Thus, investigating changes in health shortfall and socioeconomic health inequality in smoking behavior between these two years provides a valuable case study for assessing the errors associated with the linear approximation of the RIF-regression approach. Given that the primary purpose of this empirical exercise is demonstrative, we will refrain from proposing policy-based recommendations. In our concluding remarks, we highlight potential directions worth delving into.

The NHIS has monitored the health outcomes of Americans since 1957. It is a cross-sectional household interview survey representative of American households and non-institutionalized individuals collected via personal household interviews. We focus on the adult population in the 2000 and 2020 public-use data for which we have information on income. As a result, sample sizes are 31,410 for 2000 and 30,007 for 2020. We use the sample adult file to extract information on cigarette consumption and use family per capita income to infer the socioeconomic rank of individuals. In this empirical illustration, we focus on decomposing temporal changes in health shortfall and socioeconomic health inequality in cigarette consumption at the national level.

4.1 Comparison of health shortfall and socioeconomic health inequality in cigarette consumption over time

We begin by evaluating the health shortfall over time. Health shortfall (and achievement) indices are designed to capture one of the health policymaker's primary objective of understanding the health distribution. They account for both the population's health variable's average level and socioeconomic inequalities in its distribution. In Figure 2 we provide a visual representation of the dominance condition for the health shortfall associated with cigarette consumption. An initial glance at this figure reveals that the absolute health concentration curve of 2000 consistently lies above its 2020 counterpart. This non-intersection implies that there isn't any rank-dependent health shortfall index where the health shortfall for cigarette consumption in 2020 surpasses that of 2000.

This result carries two key takeaways. First, it indicates that the average cigarette consumption decreased between 2000 and 2020. This reduction can be visually confirmed by observing the shift in the coordinate of the absolute health concentration curve at $q = 1$, which represents the population average level of the health shortfall variable since $C_A(1; F_{HY}) = \mu_H$.[5] Secondly, irrespective of a policymaker's structure of preferences for socio-economic health disparities, the health shortfall related to cigarette consumption has diminished between these two decades. The ability to derive these conclusions is attributed to the use of absolute health concentration curves, as their curvature provides insights into the above-mentioned dynamics.

The dominance result between the absolute health concentration curves of 2000 and 2020 indicates that there has been an improvement across all socioeconomic ranks. Despite this clear social improvement, it is still essential, from a public health standpoint, to investigate potential socioeconomic inequalities in these gains. To better understand these disparities, we delve into the evolution of socioeconomic health inequalities over time. As visualized in Figure 3, the socioeconomic health inequality dominance condition for cigarette consumption reveals that at no point does the 2020 curve fall below that of 2000. This non-intersection implies that there isn't any rank-dependent socioeconomic health inequality index that produce a higher value of socioeconomic inequality in cigarette consumption in 2000 compared to 2020. This means that even though there was a universal decline in cigarette consumption across all socioeconomic levels, the benefits were not equally reaped

[5] Note that $C_A(1; F_{HY}) = \mu_H C_R(1; F_{HY}) = \mu_H$.

by everyone. There's a discernible rise in the socioeconomic health disparity in cigarette consumption over these two decades.

4.2 Modeling the health concentration curves

Although comparing health shortfall and socioeconomic health inequalities over time offers valuable insights from a health policy perspective, it is also important to understand the underlying changes behind these results. A logical hypothesis might be the demographic evolution between 2000 and 2020, perhaps is accounting for the decline in smoking rates. To capture the importance of the demographic changes, Table 1 presents some population characteristics in terms of three demographic aspects: age, race, and education (we will control for additional demographic characteristics in the analysis). Descriptive statistics suggests that, between 2000 and 2020, there was an increase proportion of population above 50, an increase in racial diversity, and an increase in education. These changes alone, everything else being held constant, may induce a change in the absolute and relative health concentration curves.

To construct a statistical association between the potential correlates and the values of the absolute and relative health concentration curves, we need an econometric model. The purpose our empirical analysis is to showcase how to model absolute and relative health concentration curves using the approaches described in Sections 3.3 and 3.4. For the reweighting approach, we employed a straightforward parametric logit model. To address the challenges posed by multiple discrete covariates, we use Firth's (1993) penalized maximum likelihood estimation approach. For the RIF-regressions, we rely on a simple linear model. In both cases, we use age, sex, race, region of residence, education, and an indicator variable for the presence of children at home as covariates. Our reference group for comparison is childless white males with less than a high school degree residing in a Northeastern state. We use a grid of twenty-one points to estimate the absolute and relative health concentration curves and for the overall decomposition. To be systematic, we estimated the curves using a 21-point grid $q \in \{0, 0.05, 0.10, 0.15, \ldots, 0.85, 0.90, 0.95, 1\}$. For absolute concentration curve $C_A(q; F_{HY})$ we've computed twenty regressions on a grid spanning from 0.05 to 1. For the relative concentration curve $C_R(q; F_{HY})$ we estimate nineteen regressions on a grid spanning from 0.05 to 0.95.[6] In order to estimate the values of the terms $\mathrm{E}[\,H|Y = F_Y^{-1}(q)]$ in the expression of $RIF(h_i, y_i; C_A(q; F_{HY}))$ and $RIF(h_i, y_i; C_R(q; F_{HY}))$, we use a Nadaraya-Watson non-parametric estimator (Nadaraya, 1964; Watson, 1964). We use the Silverman rule of thumb (Silverman, 1986) for the selection of the bandwidth.

4.3 Overall decomposition of changes in health shortfall and in socioeconomic health inequality

In this section, we are interested in the overall structural and composition effects of the change in absolute and relative health concentration curves from 2000 and 2020. We estimate this decompostion using the methodology outlined in Section 3.3 on our dataset. In addition to the estimated curves for both years, Figures 2 and 3 displays the estimated counterfactual curves $C_{A\langle 2000|2020\rangle}(q)$ and $C_{R\langle 2000|2020\rangle}(q)$. We use the two estimated and counterfactual curves for both the absolute and relative health concentration curves to determine the structural and composition effect in the decomposition.

The curves in Figure 4 represent the structural and composition effects of the differences in the absolute health concentration curves between 2020 and 2000. A visual inspection of Figure 4 indicates that both structural and composition effects play a role in the change between these two years. The result indicates that although both the composition and structural effects play quantitatively an equivalent role in the reduction in the shortfall in cigarette smoking at low socioeconomic ranks. However, for higher socioeconomic ranks, the dominant factor is the structural effect. Notably, for every socioeconomic rank the structural

[6] Note that, since by construction $C_R(1) = 1$, there is no need to model the health concentration curve at this socioeconomic rank.

and composition effects are negative. This suggests a consistent reduction in health shortfall due to these two effects, regardless of the specific health shortfall index used.

Figure 5 presents the structural and composition effects of the difference between the relative health concentration curves of 2020 and 2000. The effects appear to counteract each other. The changes in the distribution of demographic characteristics would have reduced socioeconomic health inequality in cigarette consumption for any rank-dependent socioeconomic health inequality index. Yet, the structural effect, which goes in the opposite direction, overshadow this composition effect accounting for the observed increase in socioeconomic health inequality in cigarette consumption between the two years.

4.4 Assessing the impact of specific demographic variables

In this Section, we investigate individual covariates' structural and composition effects. As explained in Section 3.4, we estimate these effects using RIF of the health concentration curves. Before we proceed with our estimation, let us assess how well the RIFs approximate the health concentration curves. Figure 6 displays the curves derived from averaging $RIF(C_A(q))$ and $RIF(C_R(q))$ and the estimated absolute and relative health concentration curves for both years. It illustrates that taking the average of the RIF gives the value of the health concentration curves. Nevertheless, the inherent bias in non-parametric estimation of the term $\mathrm{E}[\,H|Y = F_Y^{-1}(q)]$ in the expression of the RIF of both the absolute and relative health concentration curve, means that the point estimates are not precisely equal. Despite this inherent bias, we see from the curves in Figure 6 that averaging the RIF gives a very good approximation of the estimated curve. We opted not to include the confidence band on this figure to focus the point estimates. Nonetheless, a glance at the confidence bands in Figures 2 and 3, confirms that averaging the RIF yields clearly to a curve within these bands.

4.4.1 Estimating the RIF-regression models

We begin by estimating the six regression models we use in the analysis: the model of the 2020 health concentration curves, $C_{A\langle 2020|2020\rangle}(q)$ and $C_{R\langle 2020|2020\rangle}(q)$, the model of the 2000 health concentration curves, $C_{A\langle 2000|2000\rangle}(q)$ and $C_{R\langle 2000|2000\rangle}(q)$, and the model of the counterfactual health concentration curves of the distribution of covariates of 2020 with the returns of 2000, $C_{A\langle 2000|2020\rangle}(q)$ and $C_{R\langle 2000|2020\rangle}(q)$.

Tables 2, 3, and 4 display the estimated values of the parameters of the models of $C_{A\langle 2000|2000\rangle}(q)$, $C_{A\langle 2020|2020\rangle}(q)$ and $C_{A\langle 2000|2020\rangle}(q)$ for $q \in \{0.25, 0.50, 0.75, 1\}$. For the year 2000, nearly all covariates significantly impacted the four coordinates of the $C_A(q; F_{HY})$ with two notable exceptions. First, living in the Midwest does not significantly impact $C_{A\langle 2000|2000\rangle}(q)$ at $q = 0.25$ and 0.50. Second, having children does not significantly impact $C_{A\langle 2000|2000\rangle}(q)$ at $q = 0.75$ and 1 values. In 2020, the age covariate became statistically insignificant. In contrast, residing in the Midwest is now significant for all values of q in the table. The variable regarding the presence of children is statistically significant at q values of 0.75 and 1 but not for $q = 0.25$. Upon examining the counterfactual curve model $C_{A\langle 2000|2020\rangle}(q)$, the most significant covariates mirror those of $C_{A\langle 2000|2000\rangle}(q)$ except for the presence of children, which only holds significance at $q = 1$.

Tables 5, 6, and 7 display the estimated values of the parameters of the models of $C_{R\langle 2000|2000\rangle}(q)$, $C_{R\langle 2020|2020\rangle}(q)$ and $C_{R\langle 2000|2020\rangle}(q)$ for $q \in \{0.25, 0.50, 0.75\}$. The Hispanic and Black variables are significant for all three q values in both specified years and the counterfactual model. Being a female does not significantly impact the relative health concentration curve for all three values of q in both years and the counterfactual model. The same holds for being in the South. The significance of all other variables changes when moving from one model to the other.

4.4.2 Assessing the impact of the error in estimation

In Section 3.4, we highlighted that the main advantage of adopting a RIF-regression approach lies on the identification each covariate's effect. However, it has a limitation because it represents a linear approximation of our two curves, which are inherently non-linear functionals. Within the same section, we introduce a methodology from Firpo, Fortin, and Lemieux (2018) designed to capture the errors associated with this approximation. To assess the influence of these errors on our analysis, Figure 7 compares the structural and composition effects on the absolute health concentration curves obtained using the RIF-regression approach with the one obtained using the reweighting approach. A closer look at this figure reveals that the structural effects of both approaches are almost the same. Any error stemming from the specification of the reweighting function seems insignificant. The information on the four points in Table 8 supports this conclusion since the errors linked with the reweighting function's misspecification are insignificant. The composition effects obtained using both methods are closely aligned. The curve representing the error due to departure from linearity seems consistently positive, suggesting that the RIF-regression approach may overestimate the composition effect. Referring to the information displayed in Table 8, the relative magnitude of this error seems to be between 11% and 16.5% of the composition effect.

Figure 8 compares the structural and composition effects on the relative health concentration curves obtained using the RIF-regression approach with the one obtained using the reweighting approach. The visual inspection of the figure suggests again that the structural effects of both approaches are almost the same. Furthermore, the error due to the specification of the reweighting function seems insignificant. The information for two of the three points in Table 9, reinforces this information and confirms the insignificance of the error linked with the reweighting function's misspecification. However, a significant error is noted for $q = 0.75$ but it constitutes less than 1% of the total structural effect at that point. The composition effects estimated with the two approaches are in the same direction. However, they are less close than in the case of the absolute concentration curve. The error due to departure from linearity suggests that the RIF-regression approach overestimates the composition effect with Table 9 suggesting an overestimation ranging from 18% and 48% of the total composition effect.

In summary, the RIF-regression approach fairly precisely estimates the structural effect for both absolute and relative health concentration curves. Such accuracy is crucial, especially when analysts often focus on the structural effect, which represents the impact of assignment to group 1. This structural effect might correspond to a treatment effect or policy influence in various contexts (but not in the case of our empirical application). While the composition effect estimation is not as precise as the structural effect, the method allows for measuring errors resulting from deviations from linearity.[7]

4.4.3 Assessing the impact of individual covariates on the structural effects

Figure 9 illustrates the structural effect of each variable on the absolute health concentration curve. A visual examination reveals that the change in returns to age and gender would contribute to a robust increase in the health shortfall. However, the structural effect of the intercept, i.e., the social change that affects everyone, goes in the opposite direction, and its magnitude offsets all the other effects. This result means this social change dominates the overall structural effect on the absolute health concentration curve. Table 10 presents the estimates of the structural effect associated with each demographic characteristic. It is interesting to note that although not clear from visual inspection, the structural effects of other characteristics, even if small, remain significant. This statistical significance holds for all race-associated variables. For other variables, such as the presence of children in

[7]Although we do not perform this exercise in this paper, one could potentially allocate a portion of the error to each of the covariates using equation (19).

the household and some education variables, the structural impacts are significant for some values of q.

The structural effect of the return to age is especially noteworthy. Given that all the estimated age parameters in Tables 2 and 4 are all significant and negative, the change in the age structure displayed in Table 1 would have induced a decrease in the $C_A(q)$. However, the parameters estimated in Table 3 are not statistically significant. All this explains that the structural effect of age, which is induced by the difference in the parameters in Tables 3 and 4, would have increased health shortfall for any health shortfall indices.

Even though they are of a smaller magnitude, the structural effect of both gender and race would have induced a robust increase in health shortfall for cigarette consumption. These two effects are consistent with the change of the values of the parameters for these variables between Tables 2, 3, and 4. Specifically, there is a noticeable reduction in the magnitude of these negative parameters between 2020 on one side, and 2000 and the counterfactual of 2020 with the 2000 returns on the other side.

Figure 10 displays the structural effect of each variable on the relative health concentration curve. We note that the only demographic characteristic with a discernible structural effect is having a High School degree. As corroborated by Table 11, the other educational variables also significantly impact the relative health concentration curve in the same direction. However, these impacts are smaller in magnitude. The change in the return to education would yield a robust decrease in socioeconomic health inequality in cigarette smoking. Other variables such as age, race, and the presence of children also have minor significant impacts. However, the structural effect of the intercept, i.e., the social change that affects everyone, goes in the opposite direction. Its extensive magnitude outweighs the effect of the change in the return of education, indicating that the general social change amplifies the structural effect on the health concentration curve, increasing socioeconomic health inequality.

4.4.4 Assessing the impact covariates on the composition effects

Figure 11 illustrates the impact of the change in the distribution of each demographic variable. The change in the composition of demographic characteristics that have the most important impacts are the proportion of individuals from Hispanic backgrounds and those with education levels higher than a High School degree. All these impacts reduce the health shortfall unambiguously (i.e., for any rank-dependent index). Conversely, a rise in the proportion of individuals with only a High School education would increase the health shortfall. However, since most of this demographic shift gravitates towards higher education levels, the aggregate effect of educational changes in the population's education decreases the health shortfall. Additionally, the change in the age composition also reduces the health shortfall. Table 11 also reveals a significant change in the composition for almost all the other variables, except for the proportion of Afro-Americans, which is relatively constant over this two-decade span.

Figure 12 displays the impact of the change in the distribution of each demographic variable on the relative health concentration curve. The changes in the age distribution, the presence of children in the household, the proportion of Hispanics, other Non-Whites, and people having an associate degree would have all yielded a reduction of socioeconomic health inequalities. The change in the proportion of people with High School or Graduate degrees increases socioeconomic health inequalities. This result is also confirmed by the estimated values displayed in Table 13.

5 Conclusion

Health concentration curves have been widely used in the literature for graphically representing socioeconomic health inequality. Moreover, these curves are used as tools when testing for dominance conditions. Specifically, when these curves do not intersect, they provide a ranking for socioeconomic health inequality between two populations that is valid for any rank-dependent socioeconomic health inequality index. Absolute health concentration curves serve a similar purpose in illustrating health achievement and shortfall. However, applied researchers often aim to compare an existing joint distribution of health and income with a counterfactual for which no data is available. In such instances, it is crucial to have an econometric model for both the relative and absolute health concentration curve.

This paper leverages the RIF-regression approach introduced by Firpo, Fortin, and Lemieux (2009) to propose a novel method for modeling relative and absolute health concentration curves. By deriving the expressions of the RIF for these curves, the paper equips applied economists with the ability to employ conventional econometric methods to construct models of relative and absolute health concentration curves.

In addition, the paper offers an empirical demonstration of the proposed modeling approach by decomposing the changes in relative and absolute concentration curves of cigarette consumption in the US between 2000 and 2020. This empirical application not only demonstrates the application of this modeling in real-world health economics analysis but also highlights the process of estimating errors that arise due departures from linearity, especially in instances where there's been a significant shift in the distribution of covariates. In future work, it would be interesting to apply the potential outcome framework developed in this paper in the context of natural experiments to infer the causal impact of policy interventions on the relative and absolute health concentration curves. This apparent avenue would contribute to informing policymakers aiming at designing evidence-based policies to reduce health shortfall and socioeconomic health inequality.

References

[1] Abu-Ismail, K., V. Gantner, P. Makdissi, and M. Yazbeck (2020), Socioeconomic inequalities in child malnutrition in Egypt, *Metron*, 78, 175-191.

[2] Bleichrodt, H. and van Doorslaer, E. (2006). A welfare economics foundation for health inequality measurement. *Journal of Health Economics*, 25, 945-957.

[3] Chernozhukov, V., I. Fernandez-Vál, and B. Melly (2013), Inference on counterfactual distributions, *Econometrica*, 78, 1093-1125.

[4] DiNardo J, N. Fortin, and T. Lemieux (1996), Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach, *Econometrica*, 65, 1001-1046.

[5] Elbers, C., J.O. Lanjouw, and P. Lanjouw (2003), Micro-Level Estimation of Poverty and Inequality, *Econometrica*, 71, 355-364.

[6] Elbers, C., J.O. Lanjouw, and P. Lanjouw (2005), Imputed welfare estimates in regression analysis, *Journal of Economic Geography*, 5, 101-118.

[7] 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*, Elsevier, 30, 685-694.

[8] Firpo, S. (2007), Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 75, 259-276.

[9] Firpo, S., and C. Pinto (2016). Identification and estimation of distributional impacts of interventions using changes in inequality measures. *Journal of Applied Econometrics*, 31, 457-486.

[10] Firpo, S., N.M. Fortin, and T. Lemieux (2009), Unconditional quantile regressions, *Econometrica*, 77, 953-973.

[11] Firpo, S., N.M. Fortin, and T. Lemieux (2018), Decomposing wage distributions using recentered influence function regressions, *Econometrics*, 6(2), 28.

[12] Firth, D. (1993), Bias reduction of maximum likelihood estimates, *Biometrika*, 80, 27-38.

[13] Heckley, G., H.-G. Gerdtham, and G. Kjellsson (2016), A general method for decomposing the causes of socioeconomic inequality in health, *Journal of Health Economics*, 48, 89-106.

[14] Kessels, R. and G. Erreygers (2019), A direct regression approach to decomposing socioeconomic inequality of health, *Health Economics*, 28, 884-905.

[15] Khaled, M.A., 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, 315-331.

[16] Khaled, M.A., P. Makdissi, and M. Yazbeck (2023), On absolute socioeconomic health inequality, *European Journal of Health Economics*, 24, 5-25.

[17] Koenker, R. and G. Bassett (1978), Regression Quantiles, *Econometrica*, 46, 33-50.

[18] Makdissi, P. and M. Yazbeck (2014), Measuring socioeconomic health inequalities in presence of multiple categorical information, *Journal of Health Economics*, 34, 84-95.

[19] Nadaraya, E. A. (1964), On Estimating Regression, *Theory of Probability and Its Applications*, 9, 141-142.

[20] Neyman, J.S. (1923), Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doswiadczeń polowych, *Roczniki Nauk Rolniczych*, 10, 1-51. Translated from Polish as Neyman, J.S. (1990), On the Application of Probability Theory to Agricultural Experiments. Essay on Principles, *Statistical Science*, 5, 465-472.

[21] O'Donnell, O. and T. Van Ourti (2021), Dominance analysis, in R. Cookson, S. Griffin, O.F. Norheim, A.J. Culyer (eds.), *Distributional Cost-Effectiveness Analysis : Quantifying Health Equity Impacts and Trade-Offs*, Oxford University Press.

[22] Rubin, D. (1974), Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Education Psychology*, 66, 688-701.

[23] Rothe, C. (2015), Decomposing the Composition Effect, *Journal of Business Economics and Statistics* 33, 323-337.

[24] Schechtman, E., A. Shelef, S. Yitzhaki and R. Zitikis (2008), Testing hypotheses about absolute concentration curves and marginal conditional stochastic dominance, *Econometric Theory*, 24, 1044-1062.

[25] Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

[26] Van Doorslaer, E., and T. Van Ourti (2011), Measuring inequality and inequity in health and health care, in Glied, S., Smith, P.C. (eds.), *Oxford Handbook of Health Economics*, Oxford University Press, pp. 837-869.

[27] von Mises, R. (1947), On the Asymptotic Distribution of Differentiable Statistical Functions, *Annals of Mathematical Statistics*, 18, 309-348.

[28] Wagstaff, A. (2002), Inequality aversion, health inequalities and health achievement, *Journal of Health Economics*, 21, 627-641.

[29] Wagstaff, A., P. Paci and E. Van Doorslaer (1991), On the measurement of inequalities in health, *Social Science & Medicine*, 33, 545-557.

[30] Watson, G. S. (1964), Smooth regression analysis, *Sankhyā: The Indian Journal of Statistics, Series A*, 26, 359-372.

A Proof of propositions

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 $\widetilde{F}^t_{HY} = t\cdot\delta_{h_i,y_i} + (1-t)\cdot F_{HY}$ and . The influence function of the coordinate of the absolute health concentration curve at q is given by

$$IF(h_i,y_i;C_A(q;F_{HY})) = \frac{\partial}{\partial t}C_A(q;\widetilde{F}^t_{HY})\Big|_{t=0}. \tag{20}$$

Using equation (1), equation (20) can be rewritten as

$$IF(h_i,y_i;C_A(q;F_{HY})) = \frac{\partial}{\partial t}\left\{\int_0^{\widetilde{F}_Y^{-1}(q)}\int_0^{h_{\max}} h\left[t\cdot\delta_{h_i,y_i}(h,y)+(1-t)\cdot f_{HY}(hy)\right]\mathrm{d}h\mathrm{d}y\right\}\Bigg|_{t=0}, \tag{21}$$

$$\begin{aligned} IF(h_i,y_i;C_A(q;F_{HY})) &= IF(y_i;F_Y^{-1}(q))\cdot\int_0^{h_{\max}} hf_{HY}\left(h,F^{-1}(q)\right)dh \\ &\quad+\int_0^{F_Y^{-1}(q)}\int_0^{h_{\max}} h\delta_{h_i,y_i}(h,y)\mathrm{d}h\mathrm{d}y \\ &\quad-\int_0^{F_Y^{-1}(q)}\int_0^{h_{\max}} hf_{HY}(hy)\mathrm{d}h\mathrm{d}y. \end{aligned} \tag{22}$$

Note that $\int_0^{h_{\max}} hf_{HY}\left(h,F^{-1}(q)\right)dh = \int_0^{h_{\max}} hf_{H|Y}\left(h|Y=F^{-1}(q)\right)dhf_Y\left(F^{-1}(q)\right)$. Also $\int_0^{h_{\max}} hf_{H|Y}\left(h|Y=F^{-1}(q)\right)dh = \mathrm{E}\left[H|Y=F^{-1}(q)\right]$. From Firpo, Fortin, and Lemieux (2009), we know that $IF(y_i;F_Y^{-1}(q)) = \left[p-\mathbb{1}\left(y_i\le F_Y^{-1}(q)\right)\right]/f_Y\left(F_Y^{-1}(q)\right)$. The first term on the r.h.s. of equation (22) becomes

$$IF(y_i;F_Y^{-1}(q))\cdot\int_0^{h_{\max}} hf_{HY}\left(h,F^{-1}(q)\right)dh = \left[q-\mathbb{1}\left(y_i\le F_Y^{-1}(q)\right)\right]\cdot\mathrm{E}[H|Y=F_Y^{-1}(q)]. \tag{23}$$

The second term on the r.h.s. of equation (22) 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 $y_i\le F_Y^{-1}(q)$ and equal to 0 if $y_i > F_Y^{-1}(q)$. The second term on the r.h.s. of equation (22) becomes

$$\int_0^{F_Y^{-1}(q)}\int_0^{h_{\max}} h\delta_{h_i,y_i}(h,y)\mathrm{d}h\mathrm{d}y = \mathbb{1}\left(y_i\le F_Y^{-1}(q)\right)\cdot h_i. \tag{24}$$

Using equations (23) and (24) and the fact that the last term on the r.h.s. of equation (22) is exactly the expression of the $C_A(q;F_{HY})$ in equation (1), equation (22) becomes

$$\begin{aligned} IF(h_i, y_i; C_A(q;F_{HY})) &= \left[q - \mathbb{1}\left(y_i \leq F_Y^{-1}(q)\right)\right] \cdot \mathrm{E}[H|Y = F_Y^{-1}(q)] \\ &\quad + \mathbb{1}\left(y_i \leq F_Y^{-1}(q)\right) \cdot h_i - C_A(q;F_{HY}) \end{aligned} \tag{25}$$

Reorganizing the terms of the r.h.s. of equation (25) 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. □

Proof of Corollary 1: Since $C_R(q;F_{HY}) = C_A(q;F_{HY})/\mu_H$,

$$IF(h_i, y_i; C_R(q;F_{HY})) = \frac{IF(h_i, y_i; C_A(q;F_{HY})) \cdot \mu_H - C_A(q;F_{HY}) \cdot IF(h_i, y_i; \mu_H)}{\mu_H^2}. \tag{26}$$

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 (26) yields

$$\begin{aligned} IF(h_i, y_i; C_R(q;F_{HY})) &= \frac{1}{\mu_H}\left[p - \mathbb{1}\left(y_i \leq F_Y^{-1}(q)\right)\right] \cdot \mathrm{E}[H|Y = F_Y^{-1}(q)] \\ &\quad + \mathbb{1}\left(y_i \leq F_Y^{-1}(q)\right) \cdot \frac{h_i}{\mu_H} - C_R(q;F_{HY}) \\ &\quad - C_R(q;F_{HY}) \cdot \frac{h_i}{\mu_H} + C_R(q;F_{HY}). \end{aligned} \tag{27}$$

Reorganizing the terms of the r.h.s. of equation (27) 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. □

Table 1: Demographics

	Year	
	2000	2020
Age groups:		
18 to 29	22.15%	20.80%
30 to 39	20.77%	17.01%
40 to 49	21.07%	15.88%
50 to 59	14.78%	16.44%
60 to 69	9.71%	15.36%
70 to 85	11.53%	14.51%
Race:		
White	74.21%	63.49%
Hispanic	10.54%	16.55%
Black	11.16%	11.36%
Other non-white	4.10%	8.61%
Education level:		
Less than high school degree	17.81%	11.58%
High school degree or some college	50.10%	45.63%
Associate degree	8.95%	13.06%
Bachelor degree	15.16%	18.43%
Graduate or professional school degree	7.98%	11.30%

Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 2: Model of $C_{A\langle 2000|2000\rangle}(q)$

	$RIF(C_A\,(0.25))$	$RIF(C_A\,(0.50))$	$RIF(C_A\,(0.75))$	$RIF(C_A\,(1))$
Age	-0.01433^{***}	-0.02700^{***}	-0.03776^{***}	-0.04181^{***}
	(0.00140)	(0.00198)	(0.00228)	(0.00247)
Female	-0.37232^{***}	-0.71917^{***}	-1.13921^{***}	-1.41854^{***}
	(0.05395)	(0.07664)	(0.08785)	(0.09562)
Hispanic	-2.13851^{***}	-2.98347^{***}	-3.50215^{***}	-3.65738^{***}
	(0.09246)	(0.11583)	(0.12720)	(0.13483)
Black	-1.12650^{***}	-1.79976^{***}	-2.24514^{***}	-2.45088^{***}
	(0.08588)	(0.10629)	(0.12077)	(0.12612)
Other Non-White	-0.70387^{***}	-1.24637^{***}	-1.44816^{***}	-1.39458^{***}
	(0.12276)	(0.14455)	(0.17269)	(0.18220)
Midwest	-0.02255	0.14416	0.32051^{**}	0.54137^{***}
	(0.07356)	(0.10624)	(0.12689)	(0.14043)
South	0.18204^{**}	0.41292^{***}	0.69452^{***}	0.83207^{***}
	(0.07202)	(0.09927)	(0.11244)	(0.12211)
West	-0.19358^{***}	-0.33010^{***}	-0.41021^{***}	-0.41714^{***}
	(0.07503)	(0.09970)	(0.12091)	(0.13644)
High School/Some College	-0.52715^{***}	-0.94723^{***}	-1.28871^{***}	-1.31806^{***}
	(0.10394)	(0.13608)	(0.14609)	(0.15165)
Associate Degree	-0.77189^{***}	-1.58798^{***}	-2.27432^{***}	-2.38209^{***}
	(0.11956)	(0.16398)	(0.18355)	(0.19823)
Bachelor Degree	-1.00969^{***}	-2.10113^{***}	-3.38470^{***}	-4.20417^{***}
	(0.10402)	(0.14406)	(0.15932)	(0.17151)
Graduate/Professional	-0.94821^{***}	-1.9375^{***}	-3.36322^{***}	-4.72361^{***}
	(0.10163)	(0.14306)	(0.15963)	(0.16613)
Child in household	0.20377^{***}	0.30318^{***}	0.11746	-0.04645
	(0.06206)	(0.08869)	(0.10436)	(0.10778)
Constant	2.79328^{***}	5.08729^{***}	7.25283^{***}	8.46790^{***}
	(0.15521)	(0.20970)	(0.23105)	(0.24738)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 3: Model of $C_{A\langle 2020|2020\rangle}(q)$

	$RIF(C_A\,(0.25))$	$RIF(C_A\,(0.50))$	$RIF(C_A\,(0.75))$	$RIF(C_A\,(1))$
Age	−0.00141	−0.00110	−0.00105	−0.00054
	(0.00089)	(0.00109)	(0.00125)	(0.00134)
Female	-0.17885^{***}	-0.29176^{***}	-0.35920^{***}	-0.43021^{***}
	(0.03550)	(0.04498)	(0.05234)	(0.05429)
Hispanic	-1.10473^{***}	-1.54251^{***}	-1.66538^{***}	-1.73157^{***}
	(0.05365)	(0.06017)	(0.06657)	(0.06978)
Black	-0.57932^{***}	-0.87159^{***}	-0.99433^{***}	-1.05193^{***}
	(0.05961)	(0.07040)	(0.07595)	(0.07861)
Other Non-White	-0.26550^{***}	-0.48916^{***}	-0.47391^{***}	-0.54114^{***}
	(0.06143)	(0.07047)	(0.08057)	(0.08369)
Midwest	0.11139^{**}	0.26786^{***}	0.42319^{***}	0.48913^{***}
	(0.05331)	(0.06798)	(0.08064)	(0.08612)
South	0.21552^{***}	0.37384^{***}	0.49415^{***}	0.54495^{***}
	(0.05096)	(0.06207)	(0.06941)	(0.07507)
West	-0.14092^{***}	-0.17546^{***}	-0.20129^{***}	-0.20391^{***}
	(0.04415)	(0.05395)	(0.06333)	(0.06915)
High School/Some College	-0.53582^{***}	-0.88248^{***}	-0.98831^{***}	-0.99076^{***}
	(0.09468)	(0.11101)	(0.11768)	(0.12147)
Associate Degree	-0.69005^{***}	-1.12740^{***}	-1.39823^{***}	-1.49388^{***}
	(0.09757)	(0.12080)	(0.13140)	(0.13528)
Bachelor Degree	-0.87909^{***}	-1.60994^{***}	-2.20374^{***}	-2.44102^{***}
	(0.09359)	(0.11213)	(0.12029)	(0.12414)
Graduate/Professional	-0.83789^{***}	-1.57863^{***}	-2.29296^{***}	-2.65583^{***}
	(0.09184)	(0.11382)	(0.12099)	(0.12414)
Child in household	0.02794	0.13120^{**}	0.16158^{***}	0.14629^{**}
	(0.04351)	(0.05479)	(0.06042)	(0.06076)
Constant	1.48807^{***}	2.33986^{***}	2.88154^{***}	3.14008^{***}
	(0.12886)	(0.14907)	(0.16176)	(0.17105)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 4: Model of $C_{A\langle 2000|2020\rangle}(q)$

	$RIF(C_A\,(0.25))$	$RIF(C_A\,(0.50))$	$RIF(C_A\,(0.75))$	$RIF(C_A\,(1))$
Age	-0.01329^{***}	-0.02431^{***}	-0.03492^{***}	-0.03808^{***}
	(0.00119)	(0.00174)	(0.00202)	(0.00220)
Female	-0.34166^{***}	-0.58155^{***}	-0.95011^{***}	-1.22022^{***}
	(0.04855)	(0.06647)	(0.07922)	(0.08887)
Hispanic	-1.46619^{***}	-2.10174^{***}	-2.52646^{***}	-2.66313^{***}
	(0.07612)	(0.10479)	(0.12498)	(0.14582)
Black	-0.78948^{***}	-1.27017^{***}	-1.65045^{***}	-1.89294^{***}
	(0.07519)	(0.09656)	(0.10951)	(0.11969)
Other Non-White	-0.50937^{***}	-0.87813^{***}	-1.04539^{***}	-1.00326^{***}
	(0.09790)	(0.13781)	(0.16925)	(0.18952)
Midwest	0.02897	0.11004	0.29950^{**}	0.48972^{***}
	(0.06352)	(0.09291)	(0.11672)	(0.13839)
South	0.18450^{***}	0.35732^{***}	0.53593^{***}	0.66898^{***}
	(0.06264)	(0.08573)	(0.10327)	(0.12198)
West	−0.12023	-0.27358^{***}	-0.32188^{***}	-0.36042^{***}
	(0.06665)	(0.09062)	(0.11118)	(0.13440)
High School/Some College	-0.21678^{**}	-0.40644^{***}	-0.66472^{***}	-0.65224^{***}
	(0.09579)	(0.11876)	(0.12755)	(0.13155)
Associate Degree	-0.49588^{***}	-0.94211^{***}	-1.36483^{***}	-1.40092^{***}
	(0.11112)	(0.15516)	(0.17226)	(0.18930)
Bachelor Degree	-0.73270^{***}	-1.45328^{***}	-2.51559^{***}	-3.06790^{***}
	(0.09911)	(0.13274)	(0.14831)	(0.15559)
Graduate/Professional	-0.70296^{***}	-1.36382^{***}	-2.54786^{***}	-3.60736^{***}
	(0.09559)	(0.13380)	(0.15553)	(0.16424)
Child in household	0.08088	0.04349	−0.16878	-0.27395^{***}
	(0.05748)	(0.08355)	(0.09506)	(0.09684)
Constant	2.36477^{***}	4.30044^{***}	6.30936^{***}	7.39058^{***}
	(0.13209)	(0.18114)	(0.20149)	(0.21693)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 5: Model of $C_{R\langle 2000|2000\rangle}(q)$

	$RIF(C_R\,(0.25))$	$RIF(C_R\,(0.50))$	$RIF(C_R\,(0.75))$
Age	-0.00095^{***}	-0.00133^{***}	-0.00149^{***}
	(0.00033)	(0.00037)	(0.00027)
Female	−0.00039	0.00984	−0.01082
	(0.01241)	(0.01365)	(0.01052)
Hispanic	-0.33008^{***}	-0.28991^{***}	-0.18564^{***}
	(0.01910)	(0.01912)	(0.01369)
Black	-0.13560^{***}	-0.13857^{***}	-0.09600^{***}
	(0.01840)	(0.01749)	(0.01251)
Other Non-White	-0.09471^{***}	-0.14091^{***}	-0.10227^{***}
	(0.02777)	(0.02633)	(0.02257)
Midwest	-0.04582^{**}	-0.04014^{**}	−0.02777
	(0.01792)	(0.01954)	(0.01522)
South	−0.00992	−0.00827	0.01369
	(0.01718)	(0.01927)	(0.01355)
West	−0.02360	−0.03022	−0.02418
	(0.01739)	(0.01834)	(0.01520)
High School/Some College	-0.05096^{**}	-0.06875^{***}	-0.07433^{***}
	(0.02329)	(0.02253)	(0.01467)
Associate Degree	−0.04161	-0.08965^{***}	-0.11905^{***}
	(0.02797)	(0.03005)	(0.02284)
Bachelor Degree	0.02501	0.03763	−0.03442
	(0.02271)	(0.02608)	(0.02225)
Graduate/Professional	0.08009^{***}	0.16045^{***}	0.08408^{***}
	(0.02290)	(0.02660)	(0.02569)
Child in household	0.06028^{***}	0.09154^{***}	0.04286^{***}
	(0.01443)	(0.01566)	(0.01252)
Constant	0.42318^{***}	0.69482^{***}	0.96667^{***}
	(0.03260)	(0.03458)	(0.02565)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 6: Model of $C_{R\langle 2020|2020\rangle}(q)$

	$RIF(C_R\,(0.25))$	$RIF(C_R\,(0.50))$	$RIF(C_R\,(0.75))$
Age	-0.00091^{**}	-0.00057	-0.00043
	(0.00052)	(0.00047)	(0.00031)
Female	-0.02386	-0.02109	0.00808
	(0.02069)	(0.01939)	(0.01331)
Hispanic	-0.38090^{***}	-0.35740^{***}	-0.13002^{***}
	(0.02675)	(0.02160)	(0.01472)
Black	-0.16345^{***}	-0.16865^{***}	-0.06612^{***}
	(0.03071)	(0.02519)	(0.01494)
Other Non-White	-0.06002	-0.11695^{***}	-0.00618
	(0.03271)	(0.02913)	(0.01695)
Midwest	-0.04093	-0.02328	0.00176
	(0.03136)	(0.02997)	(0.02255)
South	0.02207	0.02987	0.01874
	(0.02872)	(0.02646)	(0.01931)
West	-0.05287^{**}	-0.03751	-0.01914
	(0.02644)	(0.02519)	(0.01993)
High School/Some College	-0.14668^{***}	-0.20442^{***}	-0.10061^{***}
	(0.04924)	(0.03756)	(0.02277)
Associate Degree	-0.13394^{**}	-0.15782^{***}	-0.08364^{***}
	(0.05222)	(0.04305)	(0.02643)
Bachelor Degree	-0.03496	-0.08598^{**}	-0.07673^{***}
	(0.04729)	(0.03845)	(0.02826)
Graduate/Professional	0.04971	0.03449	-0.00599
	(0.04691)	(0.04188)	(0.03180)
Child in household	-0.01622	0.03084	0.02644^{**}
	(0.02474)	(0.02214)	(0.01301)
Constant	0.65015^{***}	0.92155^{***}	0.99359^{***}
	(0.06543)	(0.05196)	(0.03495)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 7: Model of $C_{R\langle 2000|2020\rangle}(q)$

	$RIF(C_R\,(0.25))$	$RIF(C_R\,(0.50))$	$RIF(C_R\,(0.75))$
Age	-0.00137^{***}	-0.00158^{***}	-0.00198^{***}
	(0.00036)	(0.00042)	(0.00036)
Female	−0.01513	0.01702	−0.00556
	(0.01437)	(0.01594)	(0.01353)
Hispanic	-0.28010^{***}	-0.24832^{***}	-0.16743^{***}
	(0.02329)	(0.02684)	(0.02564)
Black	-0.11245^{***}	-0.09978^{***}	-0.06917^{***}
	(0.02141)	(0.02135)	(0.01651)
Other Non-White	-0.09079^{***}	-0.12316^{***}	-0.09518^{***}
	(0.02948)	(0.03389)	(0.03168)
Midwest	−0.03101	-0.04913^{**}	−0.02582
	(0.01952)	(0.02317)	(0.02138)
South	0.00733	0.00386	0.00821
	(0.01906)	(0.02185)	(0.01857)
West	−0.01109	−0.02988	−0.01579
	(0.01941)	(0.02198)	(0.02120)
High School/Some College	−0.01980	−0.02368	-0.05676^{***}
	(0.02694)	(0.02485)	(0.01585)
Associate Degree	−0.05291	-0.07456^{**}	-0.10035^{***}
	(0.03373)	(0.03773)	(0.02882)
Bachelor Degree	0.00526	0.04584	-0.05748^{**}
	(0.02832)	(0.03224)	(0.02556)
Graduate/Professional	0.06056^{**}	0.17221^{***}	0.07303^{**}
	(0.02834)	(0.03532)	(0.03434)
Child in household	0.05064^{***}	0.06351^{***}	0.01402
	(0.01660)	(0.01941)	(0.01488)
Constant	0.43679^{***}	0.68100^{***}	0.98928^{***}
	(0.03715)	(0.03980)	(0.02958)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 8: Overall decomposition (RIF approach), absolute health concentration curves

	$RIF(C_A\,(0.25))$	$RIF(C_A\,(0.50))$	$RIF(C_A\,(0.75))$	$RIF(C_A\,(1))$
Structural effect	-0.32604^{***}	-0.73113^{***}	-1.15205^{***}	-1.56963^{***}
	(0.02916)	(0.03931)	(0.04696)	(0.05286)
Error due to misspecification of the reweighting function	0.00046	0.00054	0.00125	0.00451
	(0.00185)	(0.00304)	(0.00414)	(0.00489)
Composition effect	-0.29910^{***}	-0.51221^{***}	-0.67703^{***}	-0.7669^{***}
	(0.01766)	(0.02378)	(0.02796)	(0.03137)
Error due to departure from linearity	0.04904^{***}	0.06596^{***}	0.07371^{***}	0.10145^{***}
	(0.00976)	(0.01176)	(0.01096)	(0.00886)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 9: Overall decomposition (RIF approach), relative health concentration curves

	$RIF(C_R\,(0.25))$	$RIF(C_R\,(0.50))$	$RIF(C_R\,(0.75))$
Structural effect	0.10544^{***}	0.11342^{***}	0.08252^{***}
	(0.01219)	(0.01230)	(0.00957)
Error due to misspecification of the reweighting function	−0.00022	−0.00061	-0.00075^{**}
	(0.00033)	(0.00037)	(0.00031)
Composition effect	-0.02752^{***}	-0.02913^{***}	-0.02292^{***}
	(0.00344)	(0.00372)	(0.00301)
Error due to departure from linearity	0.01323^{***}	0.00936^{***}	0.00420
	(0.00306)	(0.00348)	(0.00276)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 10: Detailed structural effect, absolute health concentration curves

	$RIF(C_A(0.25))$	$RIF(C_A(0.50))$	$RIF(C_A(0.75))$	$RIF(C_A(1))$
Age	0.56835***	1.11060***	1.62027***	1.79571***
	(0.06979)	(0.09682)	(0.11344)	(0.12275)
Female	0.08424***	0.14995***	0.30577***	0.40879***
	(0.03098)	(0.04094)	(0.04827)	(0.05281)
Hispanic	0.05982***	0.09254***	0.14250***	0.15416***
	(0.01500)	(0.01944)	(0.02280)	(0.02599)
Black	0.02387**	0.04526***	0.07451***	0.09551***
	(0.01074)	(0.01338)	(0.01488)	(0.01606)
Other Non-White	0.02100**	0.03349***	0.04920***	0.03978**
	(0.00957)	(0.01291)	(0.01580)	(0.01768)
Midwest	0.01746	0.03343	0.02620	−0.00012
	(0.01753)	(0.02441)	(0.03037)	(0.03482)
South	0.01167	0.00622	−0.01572	−0.04668
	(0.02953)	(0.03960)	(0.04634)	(0.05391)
West	−0.00491	0.02327	0.02860	0.03712
	(0.01880)	(0.02471)	(0.02994)	(0.03557)
High School/Some College	−0.14559**	−0.21723***	−0.14766	−0.15447
	(0.06305)	(0.07592)	(0.08030)	(0.08342)
Associate Degree	−0.02536	−0.02420	−0.00436	−0.01214
	(0.01986)	(0.02594)	(0.02870)	(0.03071)
Bachelor Degree	−0.02698	−0.02887	0.05747	0.11554***
	(0.02571)	(0.03223)	(0.03554)	(0.03699)
Graduate/Professional	−0.01524	−0.02427	0.02880	0.10750***
	(0.01540)	(0.02017)	(0.02259)	(0.02376)
Child in household	−0.01766	0.02926	0.11020***	0.14018***
	(0.02355)	(0.03235)	(0.03694)	(0.03782)
Constant	−0.87670***	−1.96058***	−3.42781***	−4.25050***
	(0.18967)	(0.23650)	(0.26463)	(0.28172)
Error due to misspecification of the reweighting function	0.00046	0.00054	0.00125	0.00451
	(0.00185)	(0.00304)	(0.00414)	(0.00489)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 11: Detailed structural effect, relative health concentration curves

	$RIF(C_R\,(0.25))$	$RIF(C_R\,(0.50))$	$RIF(C_R\,(0.75))$
Age	0.02232	0.04858	0.07410***
	(0.03022)	(0.03011)	(0.02292)
Female	−0.00452	−0.01972	0.00706
	(0.01302)	(0.01311)	(0.00982)
Hispanic	−0.01668***	−0.01805***	0.00619
	(0.00574)	(0.00556)	(0.00489)
Black	−0.00579	−0.00782**	0.00035
	(0.00423)	(0.00371)	(0.00255)
Other Non-White	0.00265	0.00053	0.00766**
	(0.00363)	(0.00370)	(0.00304)
Midwest	−0.00210	0.00548	0.00584
	(0.00769)	(0.00791)	(0.00662)
South	0.00554	0.00979	0.00396
	(0.01288)	(0.01302)	(0.01003)
West	−0.00991	−0.00181	−0.00080
	(0.00781)	(0.00804)	(0.00669)
High School/Some College	−0.05790**	−0.08248***	−0.02001
	(0.02606)	(0.02086)	(0.01281)
Associate Degree	−0.01058	−0.01087	0.00218
	(0.00823)	(0.00729)	(0.00516)
Bachelor Degree	−0.00741	−0.02429***	−0.00355
	(0.01030)	(0.00938)	(0.00705)
Graduate/Professional	−0.00123	−0.01556**	−0.00893
	(0.00625)	(0.00628)	(0.00540)
Child in household	−0.02230**	−0.01090	0.00414
	(0.00986)	(0.00967)	(0.00648)
Constant	0.21335***	0.24055***	0.00431
	(0.07685)	(0.06524)	(0.04678)
Error due to misspecification of the reweighting function	−0.00022	−0.00061	−0.00075**
	(0.00033)	(0.00037)	(0.00031)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 12: Detailed composition effect, absolute health concentration curves

	$RIF(C_A(0.25))$	$RIF(C_A(0.50))$	$RIF(C_A(0.75))$	$RIF(C_A(1))$
Age	−0.03737***	−0.07043***	−0.09851***	−0.10907***
	(0.00412)	(0.00623)	(0.00753)	(0.00823)
Female	0.00331**	0.00639**	0.01011**	0.01259**
	(0.00163)	(0.00302)	(0.00471)	(0.00591)
Hispanic	−0.13433***	−0.18741***	−0.21999***	−0.22974***
	(0.00839)	(0.01121)	(0.01285)	(0.01361)
Black	0.00056	0.00090	0.00112	0.00122
	(0.00287)	(0.00459)	(0.00573)	(0.00628)
Other Non-White	−0.03386***	−0.05996***	−0.06967***	−0.06709***
	(0.00601)	(0.00728)	(0.00867)	(0.00908)
Midwest	0.00095	−0.00608	−0.01352**	−0.02284***
	(0.00310)	(0.00451)	(0.00541)	(0.00612)
South	0.00247	0.00560**	0.00942***	0.01129***
	(0.00131)	(0.00225)	(0.00328)	(0.00382)
West	−0.00916**	−0.01562***	−0.01941***	−0.01974***
	(0.00357)	(0.00484)	(0.00584)	(0.00656)
High School/Some College	0.02519***	0.04526***	0.06158***	0.06298***
	(0.00541)	(0.00756)	(0.00872)	(0.00899)
Associate Degree	−0.03076***	−0.06329***	−0.09065***	−0.09494***
	(0.00518)	(0.00776)	(0.00927)	(0.00990)
Bachelor Degree	−0.03331***	−0.06933***	−0.11168***	−0.13871***
	(0.00473)	(0.00830)	(0.01192)	(0.01459)
Graduate/Professional	−0.03549***	−0.07252***	−0.12588***	−0.17679***
	(0.00448)	(0.00705)	(0.01011)	(0.01342)
Child in household	−0.01729***	−0.02572***	−0.00997	0.00394
	(0.00530)	(0.00758)	(0.00881)	(0.00913)
Error due to departure from linearity	0.04904***	0.06596***	0.07371***	0.10145***
	(0.00976)	(0.01176)	(0.01096)	(0.00886)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Table 13: Detailed composition effect, relative health concentration curves

	$RIF(C_R\,(0.25))$	$RIF(C_R\,(0.50))$	$RIF(C_R\,(0.75))$
Age	−0.00247***	−0.00347***	−0.00388***
	(0.00088)	(0.00099)	(0.00074)
Female	0.00000	−0.00009	0.00010
	(0.00011)	(0.00014)	(0.00010)
Hispanic	−0.02073***	−0.01821***	−0.01166***
	(0.00148)	(0.00144)	(0.00099)
Black	0.00007	0.00007	0.00005
	(0.00034)	(0.00035)	(0.00024)
Other Non-White	−0.00456***	−0.00678***	−0.00492***
	(0.00134)	(0.00128)	(0.00109)
Midwest	0.00193**	0.00169**	0.00117
	(0.00077)	(0.00084)	(0.00066)
South	−0.00013	−0.00011	0.00019
	(0.00025)	(0.00028)	(0.00021)
West	−0.00112	−0.00143	−0.00114
	(0.00082)	(0.00087)	(0.00072)
High School/Some College	0.00244**	0.00329***	0.00355***
	(0.00113)	(0.00111)	(0.00076)
Associate Degree	−0.00166	−0.00357***	−0.00474***
	(0.00112)	(0.00124)	(0.00097)
Bachelor Degree	0.00083	0.00124	−0.00114
	(0.00075)	(0.00087)	(0.00074)
Graduate/Professional	0.00300***	0.00601***	0.00315***
	(0.00087)	(0.00109)	(0.00100)
Child in household	−0.00511***	−0.00777***	−0.00364***
	(0.00124)	(0.00138)	(0.00107)
Error due to departure from linearity	0.01323***	0.00936***	0.0042
	(0.00306)	(0.00348)	(0.00276)

Note: $^{**}p < 0.05$; $^{***}p < 0.01$
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Figure 1: Relative health concentration curves

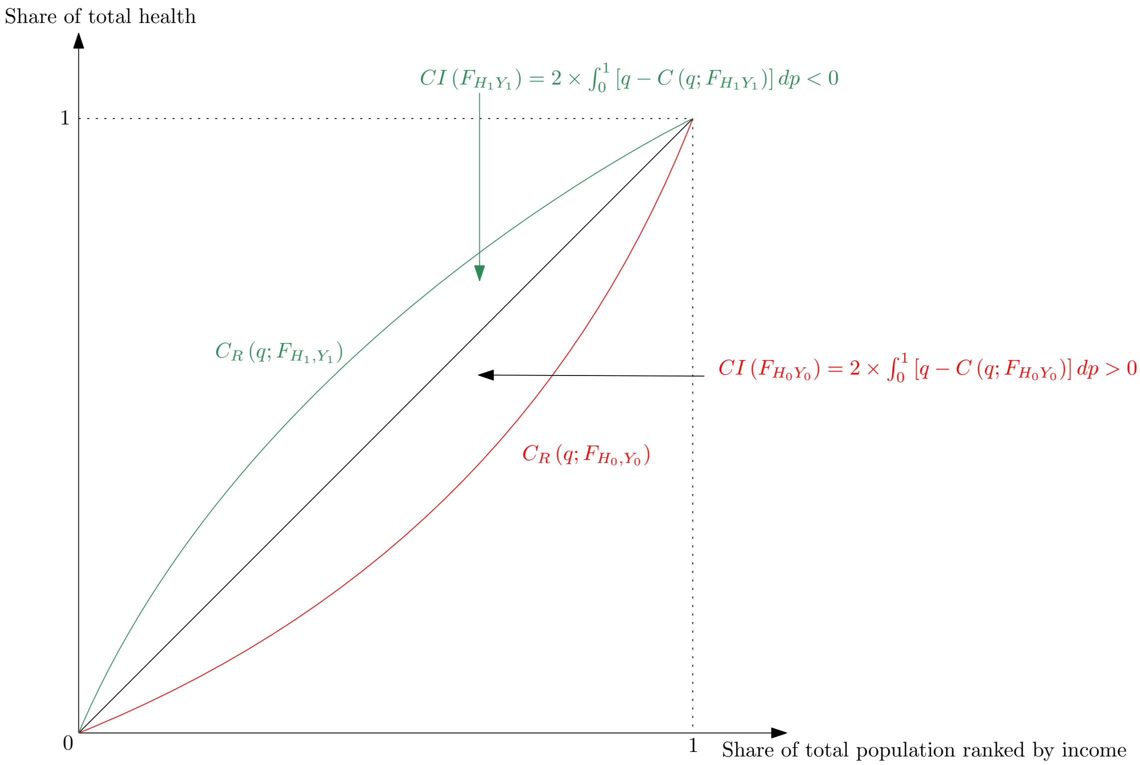
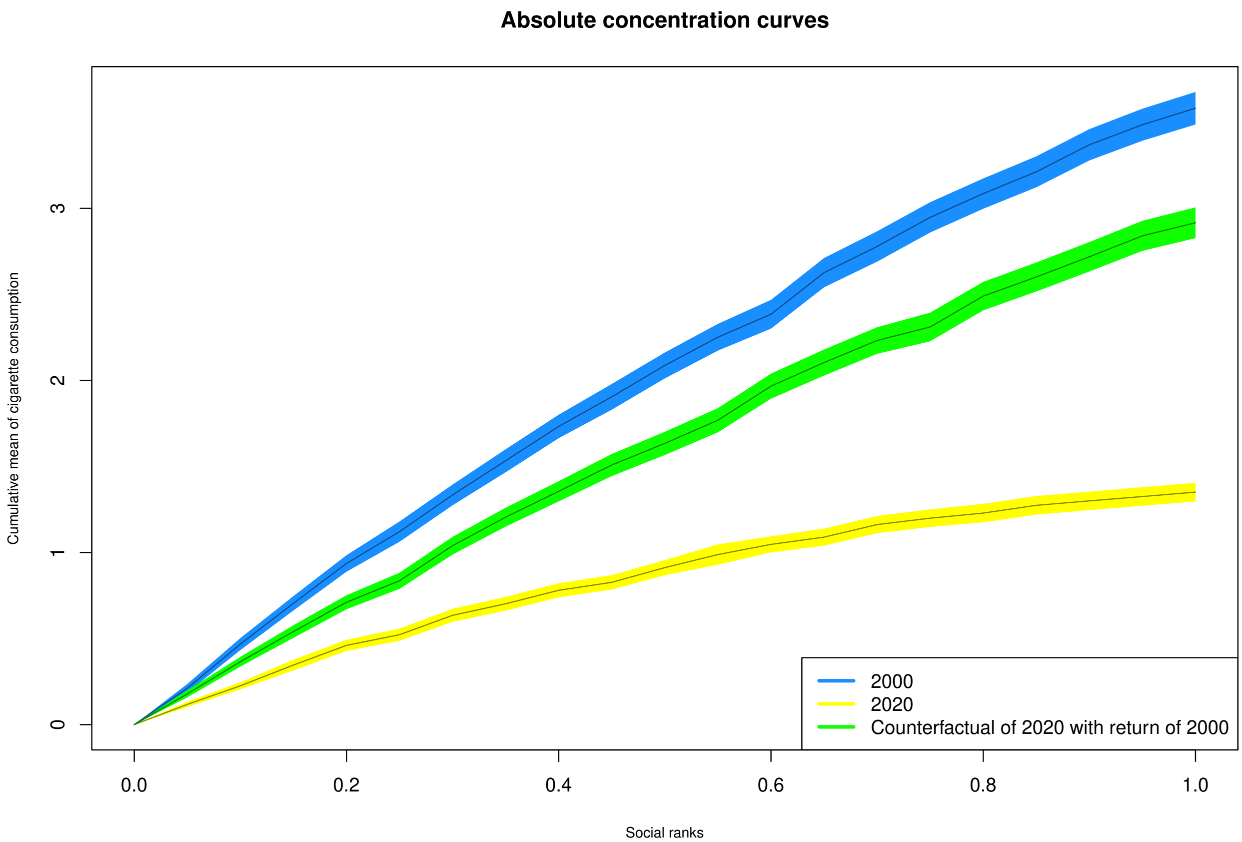
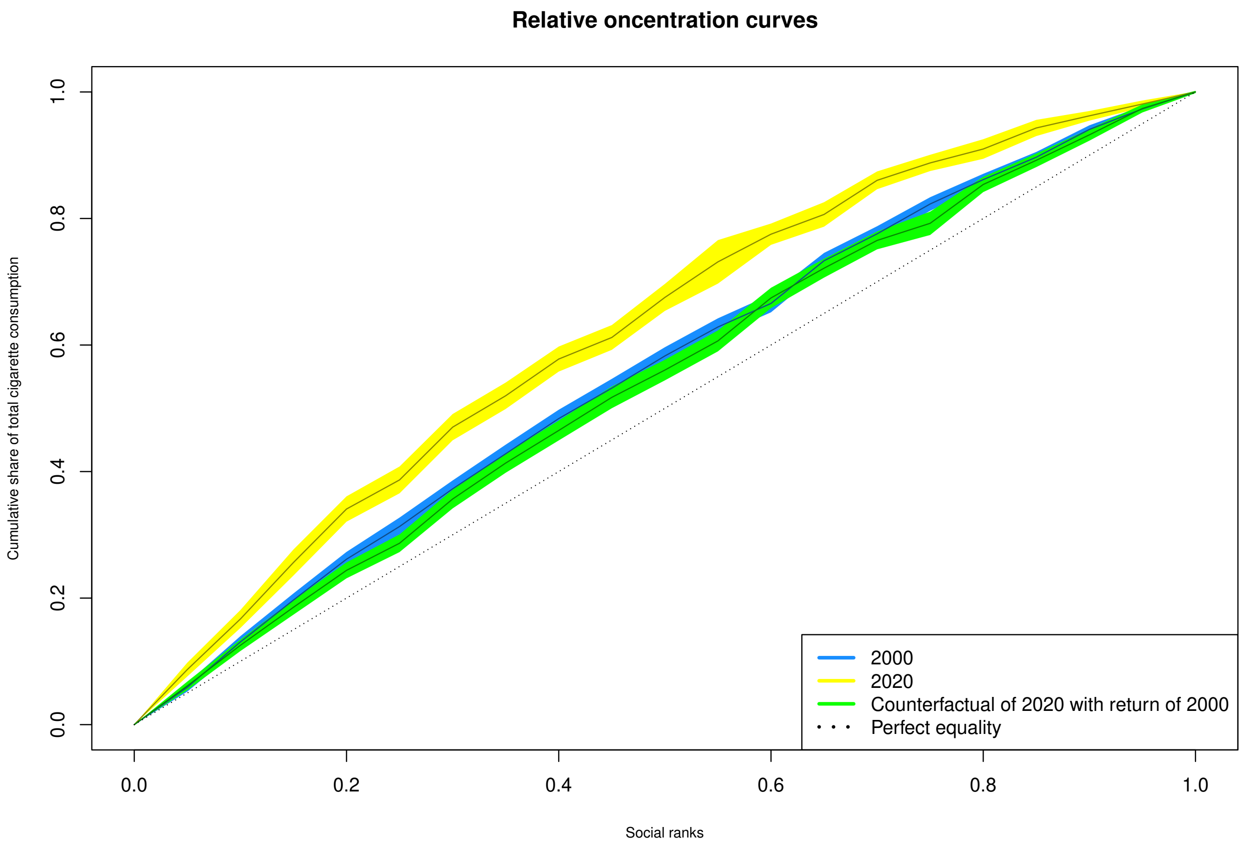


Figure 2: Health shortfall



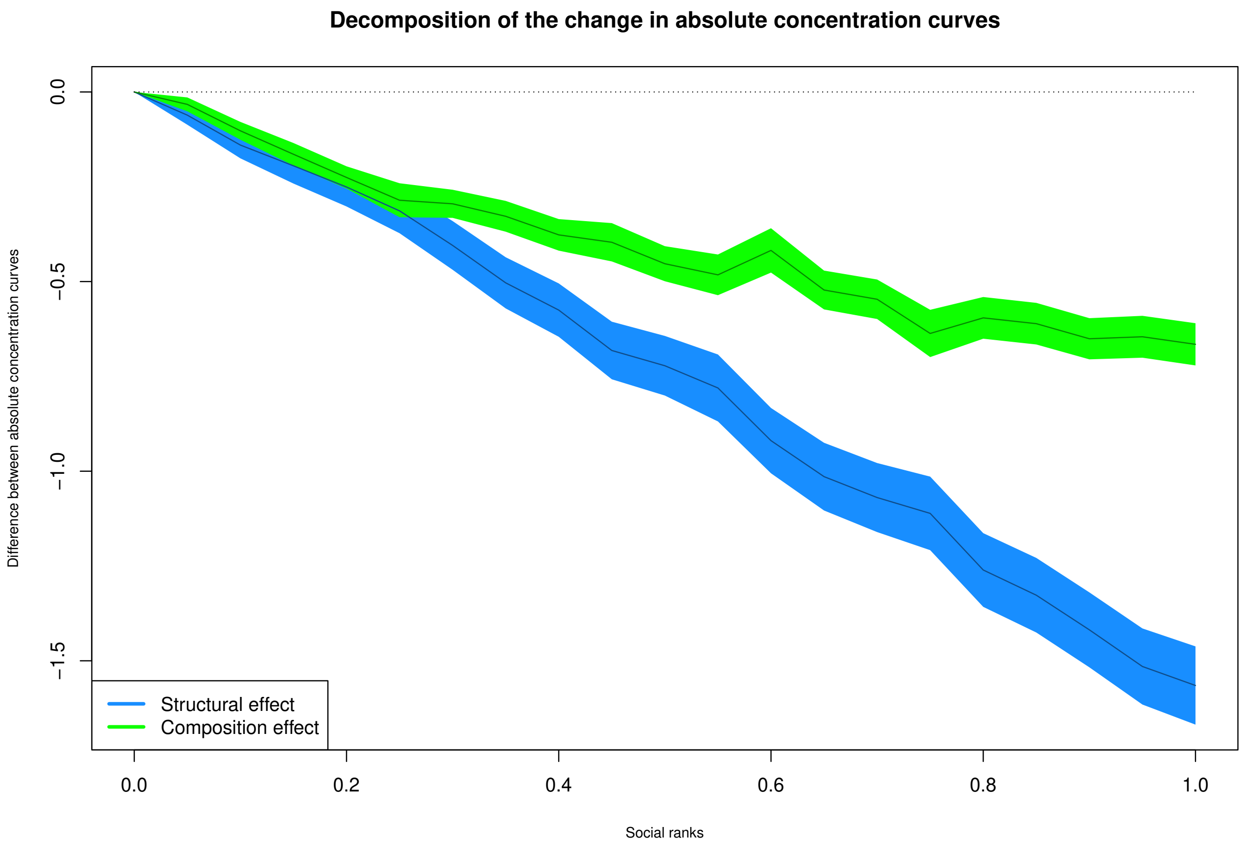
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Figure 3: Socioeconomic health inequality



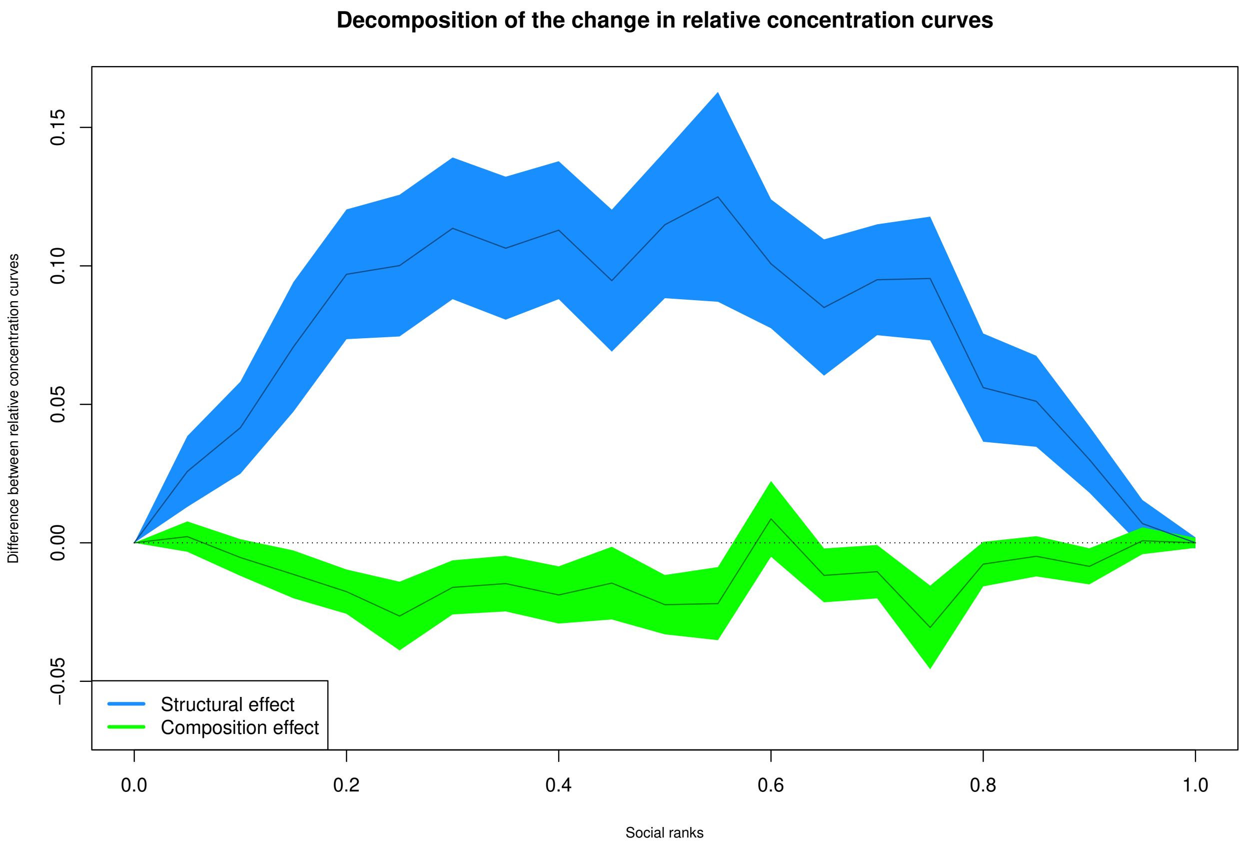
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Figure 4: Structural and endowment effects. Absolute health concentration curve.



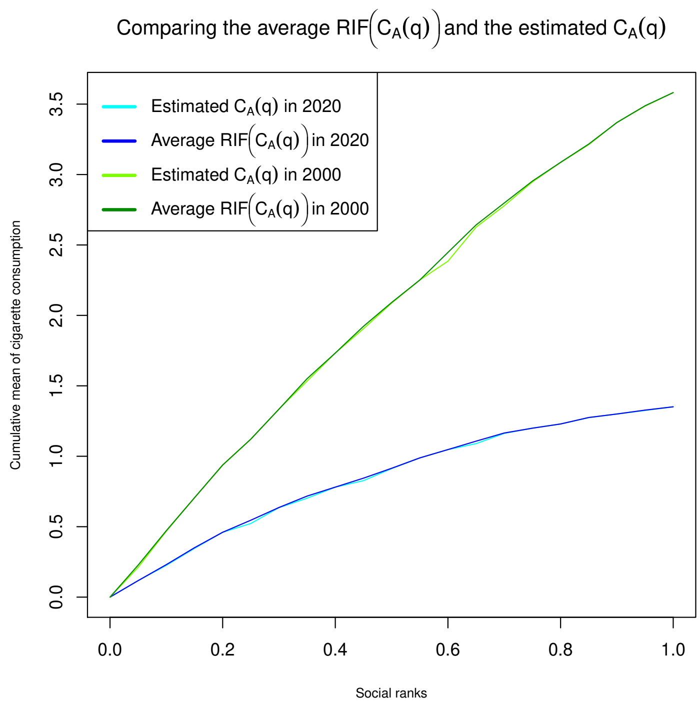
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

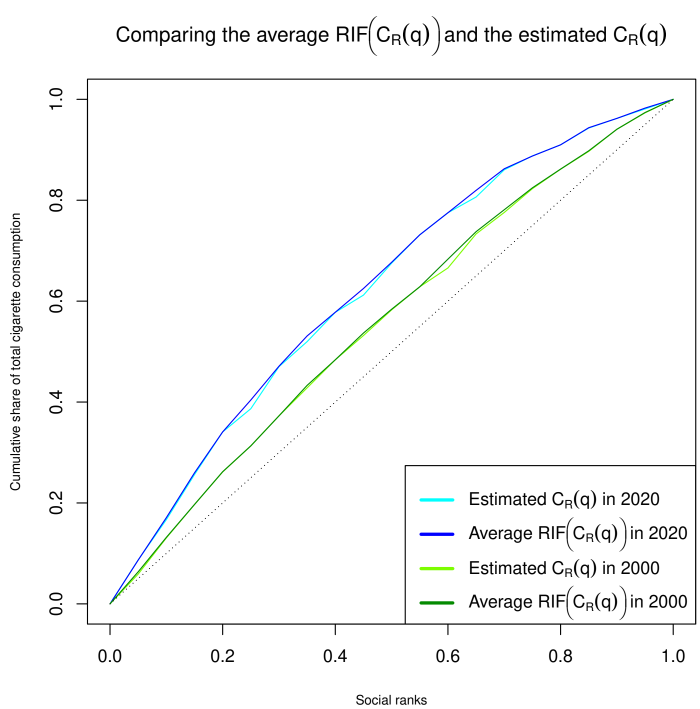
Figure 5: Structural and endowment effects. Relative health concentration curve.



Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

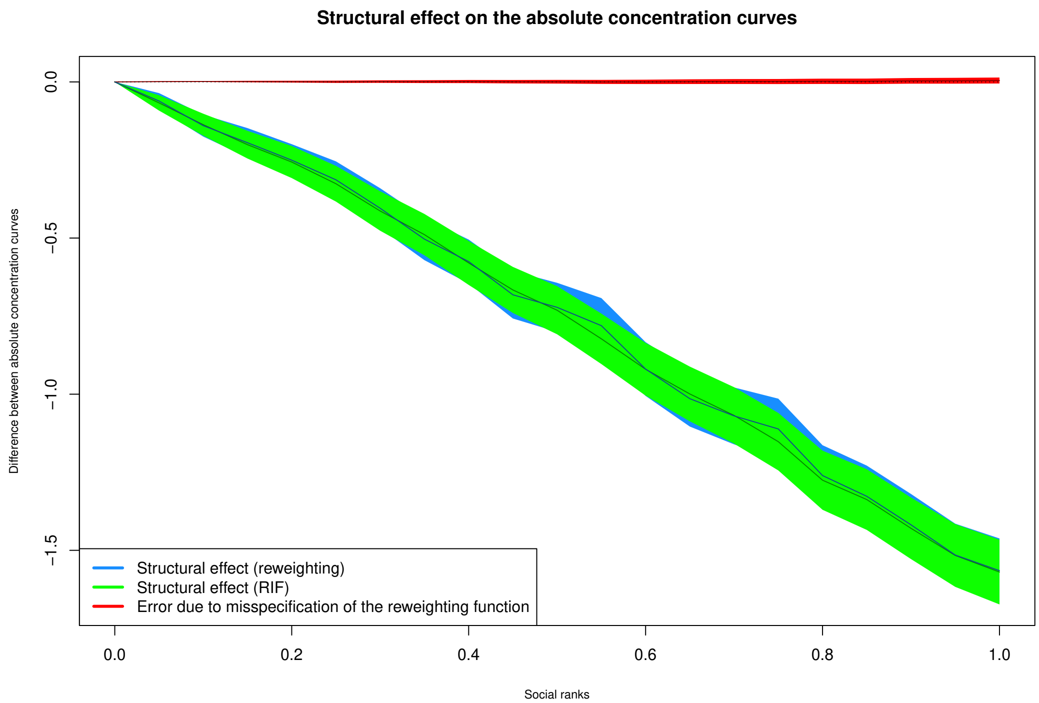
Figure 6: Illustrating the fit of the curves estimated by averaging the RIFs in the sample

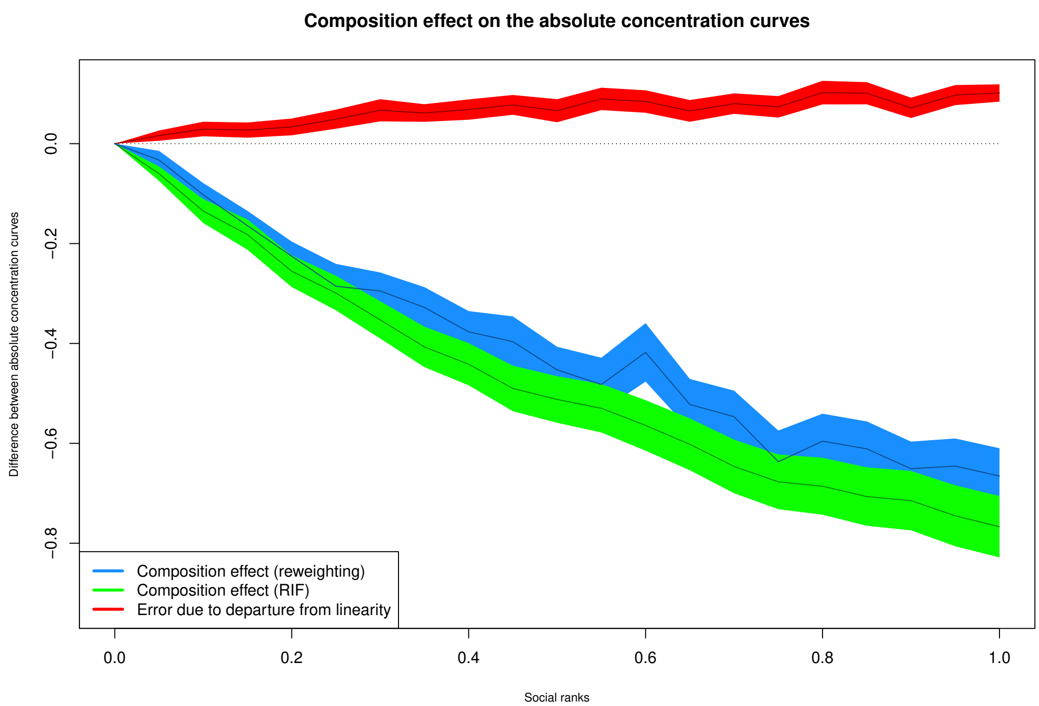




Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

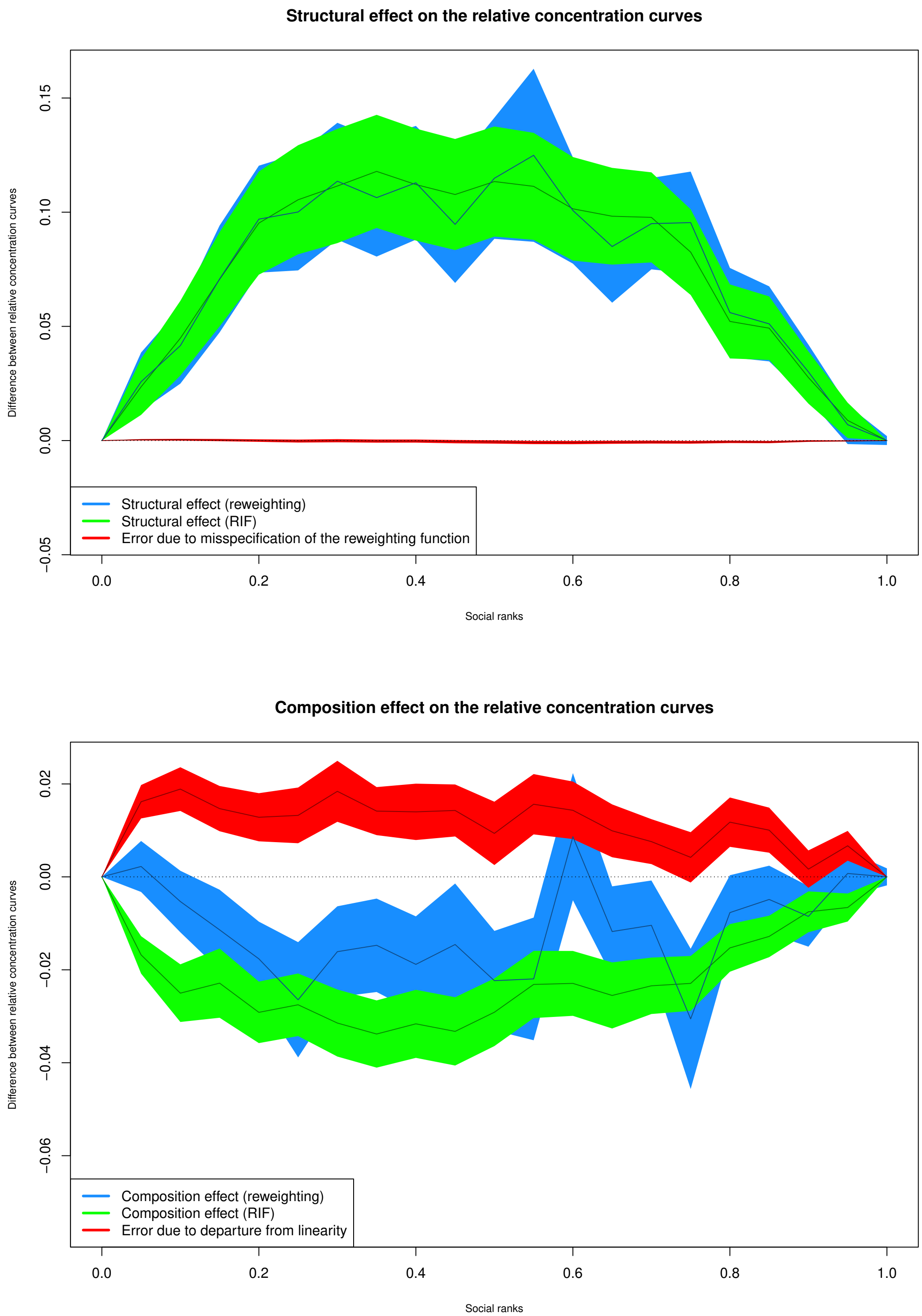
Figure 7: Comparison of the decomposition approaches. Absolute health concentration curve





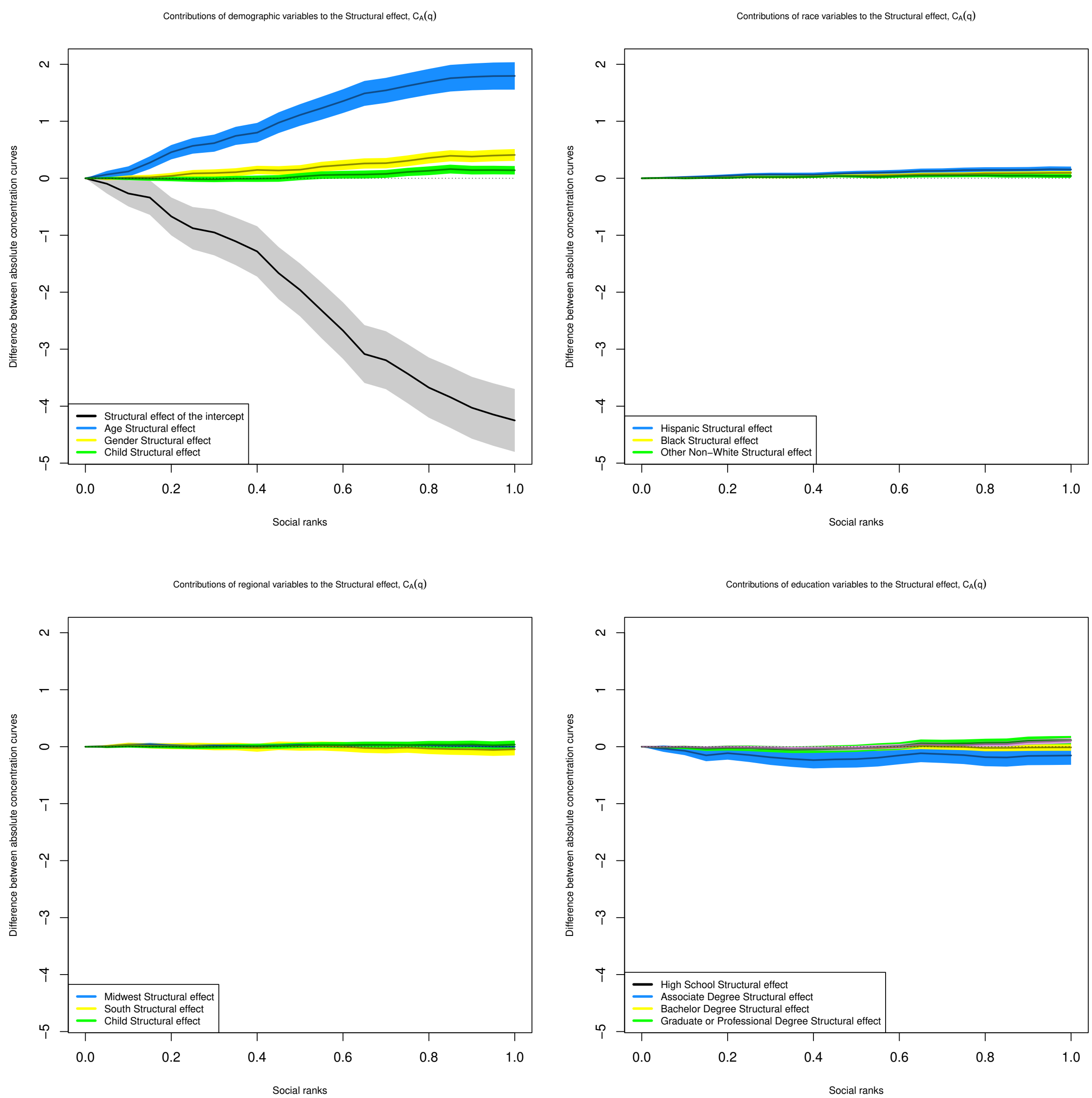
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Figure 8: Comparison of the decomposition approaches. Relative health concentration curve



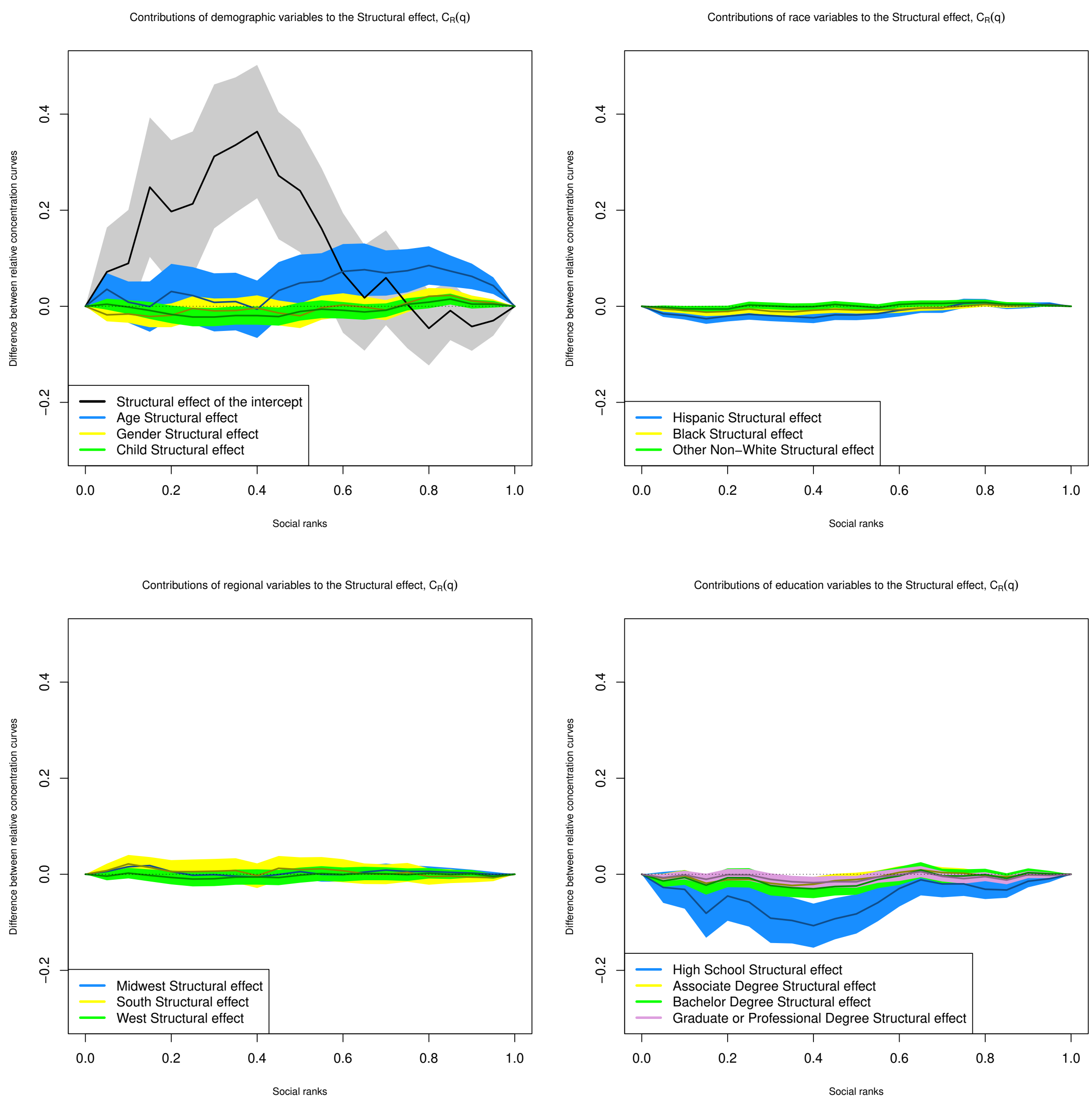
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Figure 9: Impact of each variable on the structural effect. Absolute health concentration curve



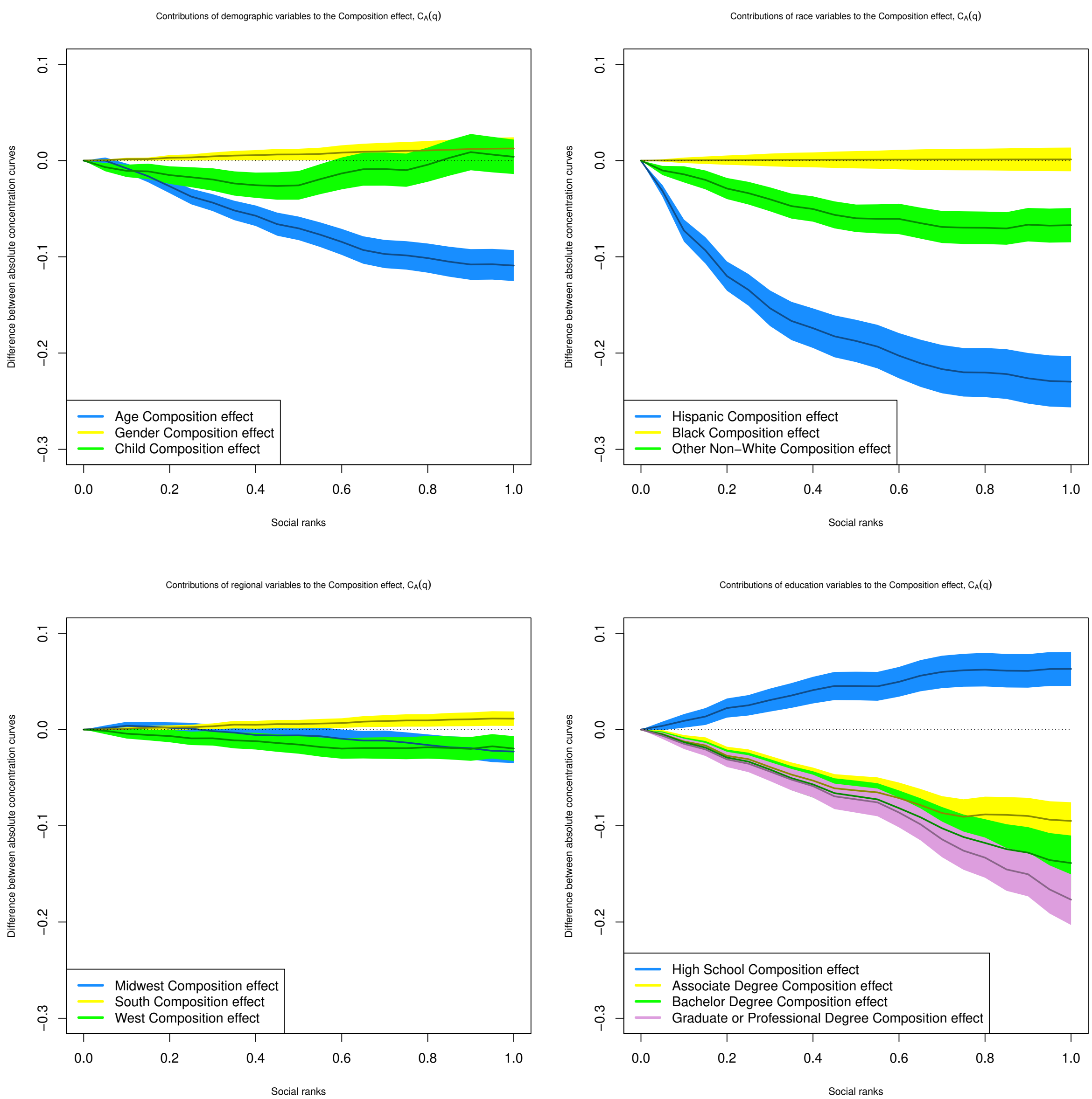
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Figure 10: Impact of each variable on the structural effect. Relative health concentration curve



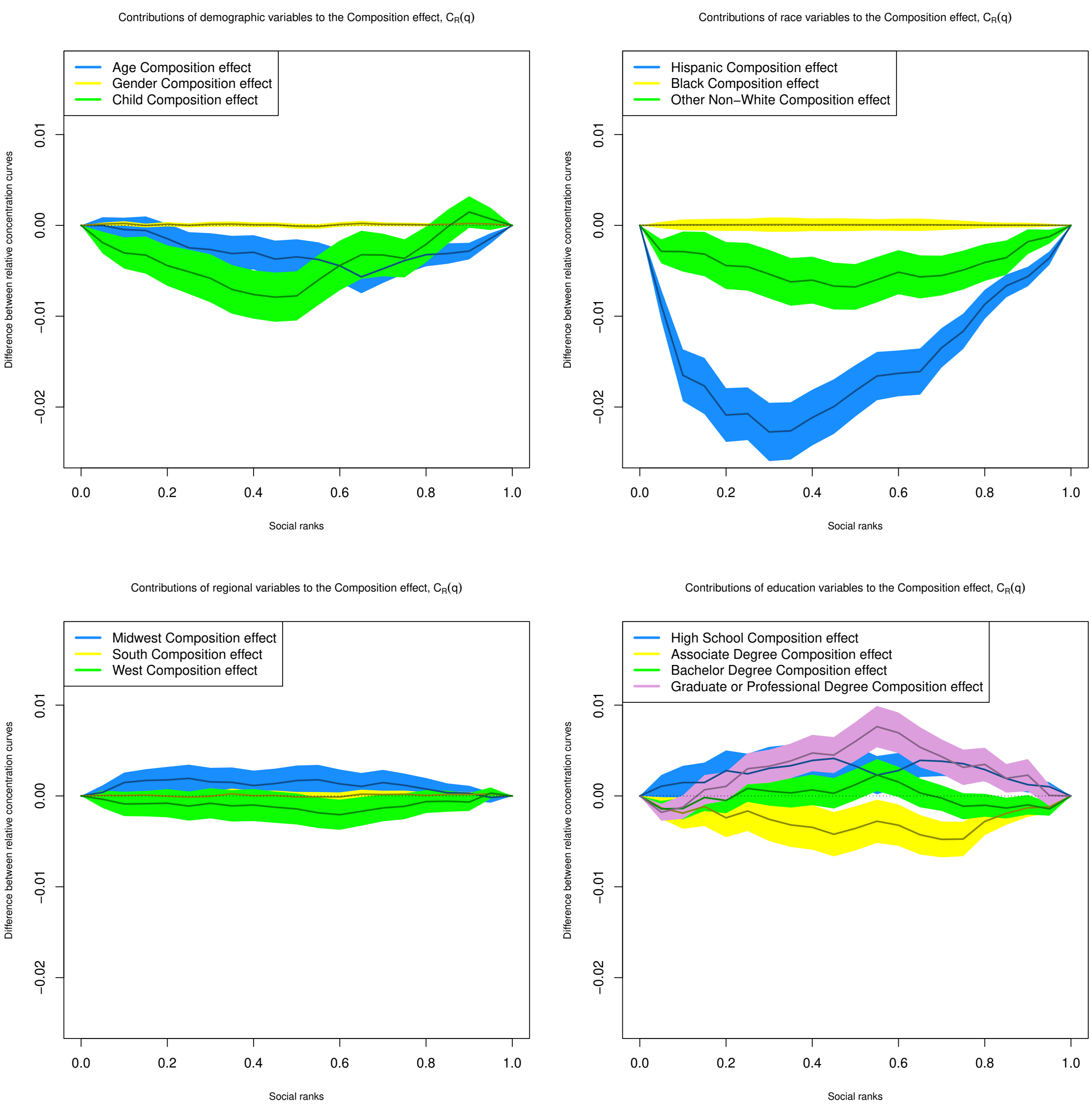
Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Figure 11: Impact of each variable on the composition effect. Absolute health concentration curve



Source: Authors' own estimation, NHIS 2000 & NHIS 2020.

Figure 12: Impact of each variable on the composition effect. Relative health concentration curve



Source: Authors' own estimation, NHIS 2000 & NHIS 2020.