HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 16/05

Inequality of opportunity in health: A decomposition-based approach

Vincenzo Carrieri & Andrew M. Jones

March 2016

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Inequality of opportunity in health: a decomposition-based approach

Vincenzo Carrieri* **Andrew M. Jones**
Università di Salerno **University of York**
HEDG, University of York **Monash University**

Abstract

This paper presents a decomposition-based approach to measure inequality in health that captures Roemer's distinction between circumstances and effort. Our approach builds on a decomposition of the Gini index with heterogeneous responses and is extended to decompose an inequality of opportunity Gini index inspired by the "fairness gap" principle. An original feature of our empirical analysis is the use of objectively measured biomarker as health outcomes and as proxies for relevant effort variables. Using data from the Health Survey for England from 2003 to 2012, we find that circumstances are the leading determinant of inequality in cholesterol, glycated haemoglobin, fibrinogen and mean arterial pressure. Moreover, we find a strong interaction between circumstances and effort leading to a smaller effect of effort on health for individuals in worse circumstances. Among the effort factors, we find that healthy diet and physical activity play the largest role in shaping objective health.

Keywords: biomarkers; decomposition analysis; health inequalities; inequality of opportunity.

JEL codes: C1, C5, D63, I14.

***Corresponding Author: Department of Economics and Statistics, Via Giovanni Paolo II, 84084 Fisciano (SA), Italy. E-mail: vcarrieri@unisa.it

1. Introduction

Evidence suggests that, at least in contemporary western liberal societies, inequalities associated with individual effort are generally considered as fair, while inequalities due to inherited factors, such as bequests or family socio-economic background, are perceived as more objectionable (Alesina and Angeletos, 2005). This reported evidence on social attitudes toward inequalities has a correspondence with a literature that has emerged in social choice theory and normative economics on equality of opportunity. Following Roemer's framework (1998, 2002), this literature separates the determinants of any outcome into two components: 'circumstances', which are not the responsibility of the individual, such as family background, gender, year of birth, and 'efforts', which to some extent are under the control of the individual. Equality of opportunity is achieved when circumstances do not play any role in the resulting outcome, which will then only depend on the exercise of individual responsibility.

Based on this framework, a number of empirical applications have dealt with the assessment of inequality of opportunity in a variety of outcomes such as income (see Ferreira and Peragine (2015) for a review) and education (Ferreira and Gignoux, 2014). The equality of opportunity principle has been advocated for the evaluation of a wide range of policies: from educational policies and their impact on health (Jones, Rice and Rosa Dias, 2011; Jones, Roemer and Rosa Dias, 2014) to policies related to the allocation of the international aid to countries for the reduction of poverty (Cogneau and Naudet, 2007).

Despite this growing interest in the concept of equality of opportunity, empirical applications remain scarce for a key determinant of human well-being: health status. This scarcity of evidence is at odds with the theoretical relevance of equality of opportunity in health advocated by many authors (i.e. Sen, 2002; Rosa Dias and Jones, 2007; Fleurbaey and Schokkaert, 2009, 2012) and with the relevance of the equality of opportunity target, which is placed at the top of the 'inequality of what' debate by many relevant institutions (e.g. World Bank, 2005).

Recently a few papers have started to deal with the measurement of inequality of opportunity in health. Rosa Dias (2009) finds considerable and persistent inequality of opportunity in health in the United Kingdom using data from the National Child Development Study. Moreover, in a follow-up study, Rosa Dias (2010) finds that inequality of opportunity in health persists accounting for the presence of unobserved heterogeneity, i.e. the partial observability of circumstances. Using data from the Survey on Health, Ageing and Retirement in Europe, Trannoy et al. (2010) find high inequality of opportunity in France according to social background and parents' longevity. Similarly, Jusot, Tubeuf and Trannoy (2013) show a high degree of equality of opportunity in France, under different normative views on the correlation between circumstances and efforts.

In this paper, we contribute to this emerging literature by proposing a decomposition-based approach to measure inequality in objective health that captures Roemer's distinction between circumstances and effort. We fully condition on circumstances by splitting our sample according to type and then estimating separate regressions of health outcomes on effort for each sub-sample. This non-parametric approach allows the model to be fully saturated in the way that it handles the circumstance variables. Using linear regression at this step generates a heterogeneous set of regression coefficients that we use in a regression-based decomposition of total inequality in the biomarkers. Note that this does not require additive separability of circumstances and effort and allows interactions between them

(through heterogeneous slopes) which is relevant for the assessment of the "fairness gap" in the spirit of Fleurbay and Schokkaert (2009, 2012).

To retrieve the relative contribution of circumstances and effort to the total inequality, we exploit a decomposition of the Gini coefficient with heterogeneous responses proposed by Jones and Lopez-Nicolas (2006) and we develop an extension of this method to complement the standard Gini with an Inequality of Opportunity Gini that measures inequality relative to the most disadvantaged type, in the spirit of the "fairness gap" principle.

Our decomposition method identifies five normatively-relevant decomposition terms: a direct and an indirect (through effort) contribution of circumstances to the total inequality, the contributions of within and between-type variation in effort to the total inequality and the contribution of randomness and luck. The between-type term is new in the equality of opportunity literature and it takes into account the contribution of the systematic variation in efforts by types on the overall inequality in health. This might be relevant in order to distinguish a "pure individual" responsibility from a "group responsibility" arising, for instance, by social contagion or social norms.

An original element of our empirical analysis is the use of biomarkers as outcome variables and as proxies of relevant effort variables such as smoking, diet and physical activity. Biomarkers are characteristics that are 'objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention'[1]. They are measured on a continuous scale associated with an increasing or decreasing risk (depending on the biomarker) of a disease state and they are often highly correlated with mortality (Rosero-Bixby and Dow, 2012; Sattar et al., 2009; Gruenewald et al., 2006). A key advantage of using biomarker data is having a measure of health which is free of reporting bias. This is particularly relevant given the possible presence of systematic reporting behaviour across individuals sharing the same set of circumstances. Indeed, previous empirical investigations show the presence of a systematic variation in reporting behaviour across socio-economic groups (e.g., Sen, 2002) which may bias the estimates of the equality of opportunity in health in a significant way.

As outcomes, we consider four biomarkers available in ten waves of the Health Survey for England (2003-2012) that are associated with some of the most prevalent diseases in all Western countries: cholesterol, glycated haemoglobin, fibrinogen and mean arterial pressure. Cholesterol measures "fat in the blood" and it is associated with a higher risk of heart disease; glycated haemoglobin is a biomarker for diabetes; fibrinogen is a haemostatic marker associated with many inflammatory diseases including cardiovascular and liver diseases and blood pressure is an important risk indicator for the development of cardiovascular diseases. Moreover, we use saliva cotinine, a major metabolite of nicotine, to objectively quantify individual smoking, and a measurement of body-mass index made by a professional nurse to properly assess the degree of individual effort related to healthy diet and physical activity. Detailed self-reported data on intensity and frequency of drinking behaviour along with information on the use of medications and equivalised income of the family complete the list of the effort variables available in our data-set. As circumstance variables, we use the cohort of birth, gender and educational level which, in combination, define eighteen Roemerian "types" that share the same circumstances.

[1] This definition is given by the National Institute of Health Biomarkers Definitions Working Group, see Atkinson et al. (2001).

We find that the target of equality of opportunity in health is still far from being reached in England. Although luck and randomness play a large role, we find that circumstances are very important in determining all of the health outcomes analysed. Moreover, we find that the strong interplay between circumstances and effort go in the direction of reducing the effect of healthy behaviours on health for individuals in worse circumstances. However, we find that individuals in worse circumstances are still empowered to reduce the risks of some specific diseases such as diabetes or inflammatory diseases. Among effort factors, we find that a healthy diet and physical activity play the largest role in determining health outcomes.

The paper is organized as follows. The next section presents the model. Section 3 introduces the data and descriptive statistics. Section 4 presents the results of our empirical analysis. The final section summarizes and concludes.

2. The Model

To model inequality of opportunity in health we adopt the framework of Roemer (2002). Roemer sorts all factors influencing individual attainment between a category of *effort factors*, for which individuals should be held partly responsible, and a category of *circumstance factors*, which, being beyond individual control, are a source of unfair differences in outcomes.

Since the outcomes of interest are health outcomes, measured by biomarkers, a generalised health production function can be defined along the lines of Roemer (2002) as $H(C, E(C))$ where C denotes individual circumstances and E denotes effort, which is itself a function of circumstances. Roemer (2002) defines social types consisting of individuals who share exposure to the same set of circumstances. The set of observed individual circumstances allows the specification of these social types in the data. A fundamental feature of this approach is the fact that the distribution of effort within each type is itself a characteristic of that type and, since this is assumed to be beyond individual responsibility, it constitutes a circumstance in itself.

The choice of circumstances and effort variables in our empirical application is largely based on the literature dealing with the measurement of inequality of opportunity in health (i.e., Rosa Dias, 2009, 2010; Jusot, Tubeuf and Trannoy, 2013). Thus, we treat as circumstances the cohort of birth, gender and educational level[2]. In the case of education, we assume that the type of secondary school in which pupils are enrolled at age 11 is beyond their individual responsibility and therefore constitutes a circumstance. This is an assumption shared by other papers (e.g. Rosa Dias, 2010). Following this strand of literature, the choice of effort variables is guided by the work on the relationship between health and lifestyles, such as Contoyannis and Jones (2004) and Balia and Jones (2008). Lifestyles are determined by the individual decisions to invest in health capital, and, therefore, they are, at least partly, within individual control. Thus, we treat as effort, cigarette smoking (saliva cotinine), alcohol frequency and intensity of consumption, dietary choices and physical activity (proxied by body mass index) and the use of medications that are, at least partly, within individual control. Moreover, we include equivalized household income as an effort variable to control for the individual's current socioeconomic position and recognise that their labour supply and productivity are, to some extent, their own responsibility (see Section 3.1 for more details).

[2] In our dataset we have detailed information on the ethinicity of the respondents. However, the inclusion of this variable among the circumstances would lead to imprecise estimates of inequality as the share of non-white individuals in the older birth cohorts is very small.

We condition on circumstances by splitting our sample according to type, τ, and estimating separate regression of health outcomes on effort for each sub-sample. This gives:

$$H_i = H\big(C_i, E_i(C_i)\big) = F_\tau(E_i) \text{ for all } \tau = 1 \ldots . T$$

Assuming linearity, we have:

$$H_i = \alpha_\tau + \beta_\tau E_i + e_i^\tau \quad (1)$$

Equation (1) gives a set of heterogeneous regression coefficients reflecting the different level of biomarkers across types (α_τ) and the different association between biomarkers and effort variables across types (β_τ), while e_i^τ is the error terms, capturing the effect of randomness or luck on biomarkers[3]. It is important to note that equation (1) does not require additive separability of circumstances and effort and allows interactions between them (through the heterogeneous slopes β_τ).

To retrieve the contribution of circumstances and efforts to total inequality we exploit the method proposed by Jones and Lopez-Nicolas (2006) who show how regression-based decomposition methods for the decomposition of health inequality, measured by the Gini index can be extended to incorporate heterogeneity in the responses of health to the explanatory variables (as in equation (1)). Moreover, we propose an extension of this method to complement the standard Gini with an Inequality of Opportunity Gini that measures inequality relative to the most disadvantaged type.

The Gini index (G) for a measure of health is given by:

$$G = \frac{2}{\overline{H}}\ Cov(H_i, R_i) \quad (2)$$

where $\overline{H} = E(H_i)$, H_i denotes the measure of health for the *i*th individual, *i* =1, ...N, and R_i denotes the cumulative proportion of the population ranked by H_i up to the *i*th individual (their 'relative rank').

Following Jones and Lopez-Nicolas (2006), we can substitute (1) into (2). This leads to:

$$\begin{gathered} G = \frac{2}{\overline{H}}\ Cov(H_i, R_i) = \\ G = \left(\tfrac{2}{N\overline{H}}\right)\textstyle\sum_i (H_i - \overline{H})\left(R_i - \tfrac{1}{2}\right) = \\ G = \left(\frac{2}{N\overline{H}}\right)\sum_i (\alpha_\tau + \beta_\tau E_i - \overline{H})\left(R_i - \frac{1}{2}\right) + \frac{2}{\overline{H}}\ Cov(e_i^\tau, R_i) = \end{gathered} \quad (3)$$

Equation (3) incorporates heterogeneity across types into the decomposition of the Gini index.

[3] In order to keep the notation simple, equation (1) is written in terms of a scalar effort variable. The extension to a vector of effort variables is straightforward and is used in our empirical application (see Jones and Lopez-Nicolas, 2006).

To provide a benchmark for our decomposition analysis, consider estimating a pooled OLS regression that treats the β_τs as constant and ignores the heterogeneity across types. In such a situation both individual and type responsibility are set equal to zero and types do not have any indirect effect (through β) in the determination of health. This implies:

$$\bar{H} = \alpha + \beta\bar{E} \tag{4}$$

Moreover, consider the OLS regressions for each type $1.....T$. Using the conditional mean from these regressions provides us with another benchmark which allows for heterogeneity in efforts between types, i.e. 'the group responsibility', and in the association between health and types, but assumes that all individuals in the same type exerted the same level of effort, thus setting 'individual responsibility' to zero:

$$\overline{H_\tau} = \alpha_\tau + \beta_\tau\overline{E_\tau} \tag{5}$$

Now, consider that:

$$(H_i - \bar{H}) = (H_i - \overline{H_\tau}) + (\overline{H_\tau} - \bar{H})$$

where:

$$(H_i - \overline{H_\tau}) = \beta_\tau(E_i - E_\tau) + e_i^\tau \tag{5a}$$

and:

$$(\overline{H_\tau} - \bar{H}) = (\alpha_\tau - \alpha) + \beta_\tau\overline{E_\tau} + \beta\bar{E}$$

Collecting some terms, $(\overline{H_\tau} - \bar{H})$ becomes:

$$(\overline{H_\tau} - \bar{H}) = (\alpha_\tau - \alpha) + (\beta_\tau - \beta)\overline{E_\tau} + \beta(\overline{E_\tau} - \bar{E}) \tag{5b}$$

By substituting (5a) and (5b) into equation (3) and changing the order of summations, the decomposition of the Gini index can be expressed as follows:

$$\begin{aligned} G = \left(\frac{2}{N\bar{H}}\right)\sum_i(\alpha_\tau - \alpha)(R_i - 1/2) + \\ + \left(\frac{2}{N\bar{H}}\right)\sum_i \overline{E_\tau}(\beta_\tau - \beta)(R_i - 1/2) + \\ + \left(\frac{2}{N\bar{H}}\right)\sum_i \beta_\tau\,(E_i - \overline{E_\tau})(R_i - 1/2) + \\ + \left(\frac{2}{N\bar{H}}\right)\sum_i \beta\ (\overline{E_\tau} - \bar{E})(R_i - 1/2) + \frac{2}{\bar{H}}\ Cov(e_i^\tau, R_i) \end{aligned} \tag{6}$$

The first term in equation (6) is the contribution to the overall inequality of the intercepts of the OLS regression across types (centred at the pooled OLS intercept coefficient). In normative terms, this measures the *direct contribution of circumstances* to the overall inequality. The second term is the covariance (weighted by the average effort level across types)

between slope parameters and the health rank. It measures the *indirect contribution of circumstances* to overall inequality, through differences in the association between efforts and outcomes across the types. The third term is the covariance between individual effort (centred at the average effort level across types and weighted by the slope parameters) and health rank. It measures the contribution of within type variation in effort on overall inequality. In normative terms, this represents the contribution of *individual responsibility* to the overall inequality. The fourth term is the covariance between the average effort by types (centred at average level of effort in the sample and weighted by the pooled OLS slope parameters) and health rank. It measures the contribution of between-type variation in effort to overall inequality and it represents the contribution of *group (type) responsibility* to the overall inequality. The final term in equation (6) is the covariance between the error term and individual rank and it measures the contribution of *randomness or luck* to overall inequality.

The Gini index in equation (2) and its decomposition can be extended in the spirit of the 'fairness gap' principle proposed by Fleurbaey and Schokkaert (2009, 2012). Indeed, the decomposition presented in equation (6) refers to a hypothetical situation in which both individual and type responsibility are set equal to zero and circumstances do not play any indirect effect in the determination of health. Another interesting benchmark scenario is represented by the health situation of the worst-off type, i.e. the group of individuals sharing exposure to the worst circumstances available in a given society. The resulting inequality index – an Inequality of Opportunity Gini - is thus expressed in terms of inequality relative the most disadvantaged type, in the spirit of the 'fairness gap' principle:

$$G^{IOp} = \frac{2}{\overline{H^w}}\ Cov(H_i, R_i) \tag{7}$$

The decomposition of G^{IOp} follows the same logic described above (i.e. in equations 3-6). The benchmark situation that ignores the type heterogeneity as in equation (4) is now replaced by the following equation for the average health of the worst-off type:

$$\overline{H^w} = \alpha^w + \beta^w \overline{E^w} \tag{8}$$

While the benchmark situation which allows for type heterogeneity is the same as equation (5). Considering that $(H_i - \overline{H^w}) = (H_i - \overline{H_\tau}) + (\overline{H_\tau} - \overline{H^w})$ and after manipulations similar to the ones shown in equations (5a) and (5b), the decomposition of G^{IOp} can be expressed as follows :

$$\begin{aligned} G^{IOp} = &\left(\frac{2}{N\overline{H^w}}\right)\sum_i (\alpha_\tau - \alpha^w)(R_i - 1/2) + \\ &+ \left(\frac{2}{N\overline{H^w}}\right)\sum_i \overline{E_\tau}(\beta_\tau - \beta^w)(R_i - 1/2) + \\ &+ \left(\frac{2}{N\overline{H^w}}\right)\sum_i \beta_\tau\ (E_i - \overline{E_\tau})(R_i - 1/2) + \\ &+ \left(\frac{2}{N\overline{H^w}}\right)\sum_i \beta^w\ (\overline{E_\tau} - \overline{E^w})(R_i - 1/2) + \frac{2}{\overline{H^w}}\ Cov(e_i^\tau, R_i) \end{aligned} \tag{9}$$

The terms in equation (9) follow the same logic of those in equation (6) but they are expressed with reference to the situation of the worst-off type. Thus, the first term gives the *direct contribution of circumstances* to the overall Inequality of Opportunity-Gini, the second term

gives the *indirect contribution of circumstances* through effort, the third term gives the *contribution of individual responsibility*, the fourth term gives the *contribution of the group (type) responsibility*, while the final term measures the *contribution of luck or randomness*. Note that the third term is virtually the same of equation (6) as it measures within type variation in efforts which do not depend on the benchmark situation chosen to perform the decomposition.

3. The Data

Our data come from the Health Survey for England (HSE). HSE is a repeated cross-sectional health interview survey of around 15,000 to 20,000 respondents conducted in England by the National Centre for Social Research (separate surveys are available for Scotland and Wales). The survey started in 1991 and has been carried out annually since then. HSE includes adults aged 16 and over, and since 1995 has also included children aged 2-15. From 2001 onwards, the survey covers all ages, but certain age groups are asked questions on selected topics only. An interview with each eligible person in the household is followed by a nurse visit for those who agree to take part. The interview includes a set of core questions, asked each year, on general health and psycho-social indicators, smoking, alcohol, demographic and socio-economic indicators, questions about use of health services and prescribed medicines. Biomarkers and health assessments are collected during nurse visits and include blood samples, anthropometric measurements, blood pressure measurements, and saliva samples. During the nurse visits, the nurse asks the respondent for permission to carry out various types of measurements. Respondents are informed about the purpose of each test and the value of each test for the monitoring of various diseases. For instance, for the cholesterol test, the nurse informs participants that "high levels are associated with blood clots, heart attack and stroke". The delivery of information is useful in order to increase compliance and establish a good working relationship.

The most popular blood-based biomarkers, which are analysed in this paper, have been collected since 2002 in the HSE. More precisely, cholesterol and glycated haemoglobin were collected from 2003 to 2012 every year; mean arterial pressure was collected from 2002 to 2012 every year, while fibrinogen was collected from 2003 to 2006 and in 2009. Other potentially relevant biomarkers (ie. tryglicerides, C-reactive protein) are collected sporadically and they are not included in this analysis. We do not make any statistical transformation of the blood-based biomarkers sample and use the valid (i.e. blood sample properly collected and successfully processed) biomarker measurements in each wave. Thus, we can use 8,979 non-missing observations for the analysis of cholesterol over the period 2003-2012, 10,065 for the analysis of glycated haemoglobin over the period 2003-2012, 10,910 for the analysis of mean arterial pressure from 2002 to 2012. The sample size for fibrinogen is much smaller: 2,336 observations from 2003 to 2006 and in 2009. For almost all the waves, blood samples are collected from individuals aged 16 and over. In a few waves a different age restriction is employed. In 2002 only individuals aged 24 or less are included; in 2004 individuals aged 11 and over are included, while in 2005 only individuals aged 65 and over are analysed. Given that we stratify by types (including birth cohorts), different age restrictions across waves are taken into account in our estimates.

3.1 Variables and descriptive statistics

In what follows, we provide a description of the variables used in our analysis. Firstly, we describe the circumstance variables that are used to define types and we discuss the

distribution of types. Then, we present the effort variables used in our analysis and the health outcomes, giving some detail on their units of measurement, the clinical cutpoints (when available) and the use of biomarker values for diagnosis of a disease state.

Circumstance variables

We use three variables to define circumstances: cohort of birth, gender and individual education. Cohort of birth is split in three categories: born before 1959; born from 1960 and 1979; born after 1979. Educational level refers to the highest academic qualification awarded and it is split in three categories: qualification below nvq3/gce a level; nvq 3/ gce a level or higher education; nvq4/nvq5/degree or equivalent. These factors reflect conditions and behaviours that are largely beyond individual control. In the case of education, we assume that the type of secondary school in which pupils are enrolled at age 11 is beyond their individual responsibility and therefore constitutes a circumstance.

A summary of these variables in our sample is presented in Table 1. Table 1 shows that around 45% of the sample were born before 1959, around 40% were born between 1960 and 1979 and around 15% were born in 1980 or later. Elderly individuals are slightly over-represented in our sample and this is due to the fact that we use only individuals with positive earnings in our analysis. The figures are nonetheless indicative of the ageing population common to many European countries. Table 1 also shows that around 45% of our sample are individuals with an education below the nvq3 level, while around the 51% of the sample are male.

On the basis of the combination of the circumstances discussed above, we can define 18 types. Type 1 is the type for which we might expect the highest penalization in terms of health outcomes: born before 1959, with a qualification below nvq3/gce a level and female. The distribution of types for each biomarker is presented in Table 2. As Table 2 shows, we have a good average sample size within each type for cholesterol, glycated haemoglobin and mean arterial pressure and, importantly, we have a relatively large sample size for the most disadvantaged type (type 1). Both aspects are relevant for our empirical analysis which is based on splitting the sample according to type and estimating separate regression of health outcomes on effort for each sub-sample. Moreover, in the case of the most disadvantaged type, a good sample size is necessary to provide an accurate benchmark for the decomposition of the Inequality of Opportunity Gini, as illustrated in Section 2. For fibrinogen, we have a much smaller average sample size (2,336 observations) and some sub-samples have few observations (i.e. types 3, 15 and 18) while the sample size of the most disadvantaged type is acceptable. For these reasons, the results related to fibrinogen should be interpreted with more caution.

Effort variables

As effort variables, we consider primarily health-related lifestyles: smoking, diet, physical activity and drinking. As a proxy of smoking, we use saliva cotinine. Cotinine is the predominant metabolite of nicotine and it is a quantitative indicator of active smoking. The key advantage of using this marker is that of having a measurement of smoking behaviour which is objective, and much less prone to the measurement errors often seen with self-reported smoking behaviour. Cotinine levels greater than or equal to 12ng/ml usually identify active smoking with high sensitivity (96.7%; Jarvis et al. 2008). As a second effort factor, we use the body mass index (BMI). BMI is a value derived from the mass (weight) and height of an individual and it is defined as the body mass divided by the square of the body height universally expressed in units of kg/m^2. Commonly accepted BMI ranges are

underweight: under 18.5, normal weight: 18.5 to 25, overweight: 25 to 30, obese: over 30. In the Health Survey for England, BMI is accurately measured by a professional during a nurse visit. Thus, measurement error issues are likely to be very small in our sample. Moreover, it represents an important risk factor for the development of cardiovascular diseases and diabetes. A high BMI might be also due to genetic factors or diseases which might be partly beyond the individual responsibility. However, both healthy diet and physical activity contribute to maintain BMI within the normal ranges. For this reason, in this paper, we use BMI as a continuous and objective proxy of both a healthy diet and of physical activity. As a proxy of drinking behaviour, we use self-reported information around the frequency of drinking during a normal week and the units of alcohol consumption on the heaviest day of the week. We take the product of these variables to take into account both the frequency and intensity of drinking of the peak of alcohol consumption.

We also include whether the individuals take medications prescribed by the doctor among the effort factors. This variable actually serves both to control for the fact that medications might be actually prescribed by the doctor in response to adverse biomarker scores and to take into account compliance with medication which is within the individual control and a matter of individual responsibility. Lastly, we consider household equivalised income which, in part, reflects an individual's effort through their labour supply and productivity and their management of household finances. This includes total income of a household from all sources, after tax and other deductions, divided by the number of household members converted into equivalised adults. In order to take into account the fact that income changes are often multiplicative in the real world (i.e. a 5% raise in wages), we use the logarithm of equivalised income in all regressions.

In Table 3, we report the mean of our effort variables by type. Average cotinine values are very heterogeneous across types. Not surprisingly, smoking behaviour is more concentrated among the most disadvantaged types. For instance, average cotinine levels are 46.53 in type 1, 120 in type 5 and just over 25 in type 18. BMI follows a less sharp pattern across types. However, the most disadvantaged types present slightly higher levels: 27.7 in type 1 vs 26.17 in type 18. Drinking behaviour is similar across types, while the use of medication seem to be much more pronounced among the most disadvantaged types. This is not surprising as these types are composed primarily by elderly individuals. Lastly, we find that more disadvantaged types tend to have lower levels of income. This is likely to be partly due to the link between education and earnings, the drop in income during retirement for the elderly and to the gender pay gap for women (the elderly, less educated and women are more represented in worst-off types).

Health outcomes

We use four blood-based biomarkers: total Cholesterol, glycated haemoglobin, fibrinogen and mean arterial pressure. As discussed in the introduction, these markers are highly predictive of some of the most prevalent non-communicable diseases. Total cholesterol (TC) is measured in units of millimoles per litre of blood, (mmol/L). The English government recommends that total cholesterol should be equal or less than 4 mmol/L among individuals at high risk of cardiovascular disease (CVD) (i.e. obese, with an history of CVD, etc.) and equal or less than 5 mmol/L or less for healthy individuals. Values above these thresholds indicate a higher risk of CVD.

Glycated haemoglobin (HbA1c) is a measure of the level of sugar in the blood over the previous 8 to 12 weeks before measurement. It is the proportion of haemoglobin proteins

that have been bound by glucose. HbA1c can be expressed as a percentage or as a value in mmol/mol. HbA1c is measured in percentages in all waves of the HSE. HbA1c values of 6.5% or more indicate diagnosis of diabetes, while values between 5.7% and 6.4% indicate pre-diabetes risk (American Diabetes Association, 2010; World Health Organisation, 2011).

Fibrinogen is a marker of inflammation and it aids the body to stop bleeding by helping blood clots to form. It is measured in grams per litre (g/L). The measure is continuous and there are no established clinical cutpoints but normal levels generally range between 1.5-3 g/L. Higher levels of fibrinogen are implicated in the development of CVD and many inflammatory diseases, such as liver diseases.

Mean arterial pressure measures the average blood pressure in an individual during a single cardiac cycle. It normally ranges between 65 and 110 mmHG. Values above this range indicate hypertension, while values below the range indicate hypotension. Mean arterial pressure is a comprehensive index of blood pressure, as it takes into account both systolic and diastolic blood pressure. Indeed, it can be easily approximated by: diastolic pressure + 1/3 (systolic pressure-diastolic pressure).

Table 4 shows the descriptive statistics of the biomarkers in our sample. We find that average biomarker values in our sample fall mostly within normal ranges, but with some exceptions. In particular, average cholesterol values are a little higher than the cutpoint of 5 while fibrinogen average scores are a little lower than the normal cutpoint of 3. Moreover, Table 1 shows higher dispersion around the average mean arterial pressure and cholesterol scores, while other biomarkers values are less dispersed around the mean.

4. Empirical Analysis

In this section, we present the results of our empirical analysis. In Section 4.1, we present the results of regressions of biomarkers on efforts. In Section 4.2, we present the results of the overall decomposition analysis while in Section 4.3 we discuss the results of the detailed decomposition by each effort variable.

4.1 Biomarker-effort regressions

In Table 5, the results of the pooled OLS regression and the OLS regression for the most disadvantaged type are reported for each biomarker. As highlighted in Section 2, these regressions are used as benchmarks on which we base the decomposition analysis of the Gini index and of the Inequality of Opportunity Gini, respectively. The complete set of regressions for all 18 types and for all biomarkers analysed is reported in Tables A.1.-A.4 in the Appendix.

Table 5 shows that effort variables generally have a significant effect on biomarkers but display a degree of heterogeneity across biomarkers and across types. With respect to the first aspect, Table 5 shows for instance that higher cotinine values are positively and significantly associated with higher glycated haemoglobin and fibrinogen levels, but they are not significantly associated with cholesterol levels. The association between cotinine and mean arterial pressure is found to be generally negative, but this depends on a different association between cotinine and diastolic and systolic blood pressure. Indeed, we find that cotinine increases diastolic blood pressure levels (i.e. "minimum" blood pressure) but it is

poorly and – in some type regressions also negatively- associated with systolic blood pressure (i.e. "maximum" blood pressure)[4]. Instead, BMI is positively associated with all biomarkers analysed. Unsurprisingly, BMI-biomarker association is steeper for mean arterial pressure and cholesterol. In the case of mean arterial pressure, a unit increase in BMI is associated with an increase of 0.67 points in the mean arterial pressure. With respect to drinking consumption, we find that it is positively associated only with higher cholesterol levels and poorly associated with the other biomarkers. Medication is generally positively associated with the biomarkers analysed. As anticipated in Section 3.1, this might be due to the fact that some individuals were already diagnosed with some diseases and in treatment for such diseases with the use of medications. Lastly, we find a large degree of heterogeneity in the income-biomarker association across different biomarkers. We find a negative income gradient for glycated haemoglobin and fibrinogen levels but a positive association between income and both cholesterol and mean arterial pressure. Similar results have been found also by Carrieri and Jones (2015) analysing the income-biomarker association on Health Survey for England data.

With respect to the heterogeneity across types, Table 5 shows that the worst-off type exhibits generally lower slope coefficients on efforts than the average. This is particularly evident in the case of BMI. For instance, the slope of cholesterol levels with respect to BMI is 0.047 points on average, and only 0.003 for the worst-off type. Similarly, BMI actually increases mean arterial pressure by 0.647 point on average and by 0.185 for the worst-off type. A similar pattern is found for cotinine, medication and income, while drinking-biomarker association seems to be more homogenous across types. The comparison of the constant terms of the pooled OLS and the OLS on the worst-off type in Table 5 reveals also a large degree of heterogeneity in the average biomarker levels across types. Heterogeneity is particularly marked in the case of cholesterol, fibrinogen and mean arterial pressure. For instance, the worst-off type presents an average cholesterol level of 5.836 which is much higher than the clinical cutpoint of 4 indicated for individuals with a higher risk of cardiovascular diseases and also higher than the normal threshold of 5. Similarly, the worst-off type presents pathologic fibrinogen levels (3.178, while the norm is 3) and a high mean arterial pressure (70.97). On the other hand, average biomarkers levels are within the normal ranges for all biomarkers, as shown by the Pooled OLS regressions. This result anticipates that circumstances play a large direct effect in influencing health outcomes.

One potential concern of the results shown above might be represented by the relatively small sub-samples by type for some biomarkers, especially fibrinogen. This might artificially increase the heterogeneity of health-efforts association across types and just capturing sampling variation in the slope coefficients across these relatively small sub-samples. In order to rule out this possibility, we experimented with the definitions of types by splitting the education in two categories (below/above NVQ3). This leads to 12 types (instead of the 18 types actually employed) and to an average sub-sample size of around 200 individuals for fibrinogen, 750 for cholesterol, 840 for glycated haemoglobin and 900 for mean arterial pressure. Our regression results are substantially unchanged. More importantly, we find that both the sign and the magnitude of the decomposition results shown in the next paragraphs are substantially confirmed under this alternative definition of types[5].

[4] Results not shown but available upon request.

[5] Results not shown but available upon request.

4.2 Overall decomposition results

The results of our decomposition analysis are reported in Tables 6-9 for cholesterol, glycated haemoglobin, fibrinogen and mean arterial pressure, respectively. In each table, we report the decomposition of the Gini index (in the top panel) and of the Inequality of Opportunity Gini (in the bottom panel) into the five contributions: individual responsibility, direct circumstances, indirect circumstances, group responsibility and a residual term. All terms are expressed in units and as percentage of the total inequality indexes. Tables 6-9 also contain the detailed decomposition of all terms into the contribution of each effort variable and these results are discussed in Section 4.3.

Tables 6-9 show that the largest contribution to the predicted inequality is represented by circumstances for all biomarkers. This is mostly due to the direct component. Indeed, the direct contribution of circumstances ranges from 5.7% of inequality for mean arterial pressure to 44% of inequality in cholesterol. The indirect contribution of circumstances is the second leading component of inequality in all biomarkers. Its contribution is negative in all analyses. As shown in Section 4.1, this implies that the types who have lower rankings in the distribution of biomarkers (i.e. worse health) have lower slope coefficients on effort. The indirect contribution of circumstances ranges from around -2% for mean arterial pressure to around -27% for cholesterol.

The third contribution to inequality is attributed to individual responsibility, i.e. within-type variation in effort. Its contribution ranges from 3.87% for cholesterol to 13% for fibrinogen. The contribution of 'group responsibility', i.e. between-type variation in effort, is the least important component of inequality for all biomarkers. Its contribution ranges from 1.3% for mean arterial pressure to 5.28% for glycated haemoglobin.

With respect to the residual term, we observe a very large contribution in the case of mean arterial pressure, amounting to around the 90% of the overall inequality. For the other biomarkers, the contribution is less important ranging from 72% to 77%. Nonetheless, this large contribution of the residual terms suggests that chance is the most important determinant of health outcomes, confirming the random and unpredictable nature of health.

The patterns described above are essentially common to all biomarkers. For mean arterial pressure only, individual responsibility seems to be slightly more important than indirect circumstances becoming the second - instead of the third - cause of inequality. Despite this minor exception, this indicates that there is a general pattern of the causes of inequality which is common to all of the health outcomes. On the other hand, the magnitude of the contributions exhibits a larger degree of heterogeneity across biomarkers. Cholesterol seems to be the biomarker mostly determined by circumstances followed by fibrinogen and glycated haemoglobin, while the role of circumstances and effort is more balanced for mean arterial pressure.

Lastly, we find that overall inequality levels are heterogeneous across biomarkers. Overall inequality is generally low for glycated haemoglobin (0.05), it is higher for fibrinogen (0.11) and cholesterol (0.12) and very large for mean arterial pressure (0.21). In contrast, the comparison of the Gini index and of the Inequality of Opportunity Gini in Tables 6-9 shows that both perspectives give a similar picture of the inequality. Estimated overall inequality is only a little smaller when measured with reference to the most disadvantaged type, in the spirit of the fairness gap principle. The discrepancy between estimated Gini and Inequality of Opportunity Gini ranges from around 4% for glycated haemoglobin to around 10% for

cholesterol, while both the magnitude of the contributions and the ranking of the causes of inequality is very similar under both indexes and for all biomarkers analysed.

4.3 Detailed decomposition results

Tables 6-9 show the detailed contribution of each effort variable to the overall inequality. We find that, in terms of *individual responsibility*, physical activity and healthy diet, proxied by BMI, play the largest role to the overall inequality across the majority of biomarkers analysed. The contribution of BMI constitutes almost 50% of the total contribution of individual responsibility to the overall inequality for all biomarkers, and, in the case of cholesterol, this actually reaches 75% of the total contribution. The contribution of the other effort variables is more heterogeneous across biomarkers. Smoking is the most important effort variable for mean arterial pressure and the second contributing factor for fibrinogen, while the use of medication is relevant for glycated haemoglobin and cholesterol. A residual role for all biomarkers is due to drinking behaviour and income.

The contribution of effort variables through the *group responsibility* terms is actually more homogenous across biomarkers. In all cases, we find that BMI and the use of medication are the most relevant contributing factors. Their joint contribution explains almost entirely the total contribution of the group responsibility terms to overall inequality.

The detailed decomposition of the *indirect circumstances* terms suggests that BMI and income play the largest role. This was also evident from the regression analysis discussed in Section 4.1 where substantial heterogeneity of the BMI and income coefficients was found across the regressions by type. Interestingly, we find some differences in the sign of this contribution across biomarkers. In the case of mean arterial pressure, cholesterol and fibrinogen, we find a negative contribution for BMI. This implies that BMI is relatively less important for health status for individuals in the lower rankings of the health distribution (i.e. the worst-off types). Conversely, in the case of glycated haemoglobin we find a positive contribution of BMI which implies a steeper slope for disadvantaged types. The contribution of income is negative for all biomarkers with the exception of mean arterial pressure. This indicates that income-biomarker association is generally lower for the worst-off types and higher for more advantaged types. This result is in line with Carrieri and Jones (2015) who found a steeper income gradient at the highest quantiles of the biomarker distribution. In normative terms, this result suggests that the hypothesis of the separability of circumstances and effort does not find support in these data and that it is relevant to allow for a heterogeneous relationship between circumstances and effort across types.

5. Conclusions

In this paper, we propose a new and relatively easy to-implement decomposition method to assess inequality of opportunity in health. The method is grounded on the theoretical framework proposed by Roemer (2002) which sorts all factors influencing individual attainment between a category of *effort factors* , for which individuals should be held partly responsible, and a category of *circumstance factors*, which are a source of unfair differences in outcomes. Our method builds on the decomposition of the Gini index with heterogeneous responses proposed by Jones and Lopez-Nicolas (2006) and it is extended to complement the standard Gini with an Inequality of Opportunity Gini index that measures inequality

relative to the most disadvantaged type, in the spirit of the "fairness gap" principle of Fleurbaey and Schokkaert (2009, 2012).

Our approach identifies the contribution of five normatively-relevant decomposition terms: direct and indirect (through effort) contributions of circumstances to the total inequality, the contributions of within and between-type variation in effort to the total inequality and the contribution of randomness and luck. Notably, this allows us to distinguish the contribution of "pure individual responsibility" (within-type variation in efforts) from the contribution of "type responsibility"(between-type variation in efforts) arising, for instance, by social contagion or social norms. Importantly, our approach does not require additive separability of circumstances and effort and allows interactions between them, through heterogeneous slopes. Moreover, it also allows a detailed decomposition of the overall terms into the contribution of each of the effort variables to the overall inequality.

We estimate inequality of opportunity in four biomarkers available in ten waves of the Health Survey for England (2003-2012) associated with some of the most prevalent non-communicable diseases : cholesterol, glycated haemoglobin, fibrinogen and mean arterial pressure. Moreover, we use Saliva cotinine, a major metabolite of nicotine, to objectively quantify individual smoking, and a measurement of body-mass index made by a professional nurse to properly assess the degree of individual effort related to healthy diet and physical activity. As effort variables, we also use detailed self-reported data on the intensity and frequency of drinking behaviour along with information on the use of medications and equivalized income of the family. As circumstances variables we use the cohort of birth, gender and educational level to build eighteen Roemerian "types".

A key advantage of using biomarkers in our analysis is having measures of health and health behaviours which are free of subjective reporting bias. This is relevant given the possible presence of systematic reporting behaviour by individuals sharing the same set of circumstances which may substantially affect the measurement of equality of opportunity in health.

Our analysis leads to a number of contributions to the existing literature. Notably, we shed light on the extent and the causes of the inequality of opportunity in a key determinant of human well-being, which is scarcely investigated by the inequality of opportunity literature. As a first result, we find that there is a general pattern in the causes of inequality which is common to all health outcomes. The largest contribution to inequality is represented by a direct effect of circumstances for all biomarkers. The indirect circumstances effects is generally the second leading causes of inequality. Its contribution is negative in all analyses which implies that the types who have lower rankings in the distribution of biomarkers (i.e. worse health) have lower slope coefficients on effort. The third cause of inequality is generally represented by the individual responsibility, i.e. within-type variation in effort, while group responsibility is found to be the least important cause of inequality for all biomarkers.

We also find that the magnitude of these contributions is heterogenous across biomarkers. The direct effect of circumstances ranges from the 5.7% of inequality of mean arterial pressure to the 44% of inequality in cholesterol. The indirect contribution of the circumstances ranges from around -2% for mean arterial pressure to around -27% for cholesterol. The contribution of individual responsibility ranges from the 3.87% for cholesterol to the 13% of fibrinogen, while the contribution of group responsibility ranges from 1.3% for mean arterial pressure to 5.28% for glycated haemoglobin. For all biomarkers we find a large contribution of luck and randomness ranging from 72% of inequality in

glycated haemoglobin to 90% of total inequality in mean arterial pressure. This confirms that luck and randomness remains the most important determinant of health outcomes.

As a third result, we find, through detailed decomposition analysis, that physical activity and healthy diet, proxied by BMI, play the largest role to the overall inequality across the majority of biomarkers analysed. The contribution of BMI constitutes almost 50% of the total contribution of individual responsibility to the overall inequality for all biomarkers, and, in the case of cholesterol, this actually reaches 75% of the total contribution. At the same time, we find a strong interplay between circumstances and effort especially in the case of BMI and income. Interestingly, we find some differences in the sign of this contribution across biomarkers. In the case of mean arterial pressure, cholesterol and fibrinogen, we find a negative contribution for BMI. This implies that BMI is relatively less important for health status for individuals in worse circumstances. Conversely, in the case of glycated haemoglobin, we find a positive contribution of the BMI which implies a higher effect on health across more disadvantaged types.

Finally, we find that both the extent and the causes of inequality are very analogous under the perspective of measuring inequality relative to the average (Gini index) or to the most disadvantaged type (Inequality of opportunity Gini). The discrepancy between overall indexes only ranges from around 4% for glycated haemoglobin to around 10% for cholesterol.

All these results suggest that the target of equality of opportunity in health, advocated by many international institutions (e.g., World Bank, 2005) is still far from being reached in England. Although randomness plays a large role, our results suggest that circumstances are still a key source of health inequalities. Moreover, our results suggest a strong interaction between circumstances and efforts in the direction of reducing the effect of healthy behaviours on health across individuals in worse circumstances. This reduces the possibility of decreasing inequalities through higher individual efforts. At the same time, we find that individuals in worse circumstances are still empowered to reduce the risks for some specific diseases. In particular, the adoption of a healthy diet and physical activity seem to be highly important to reduce the risk of diabetes. Moreover, factors partly within individual control, i.e. unhealthy diet, lack of physical activity and smoking, play a prominent role in determining the risk of many inflammatory diseases leading to high fibrinogen levels.

Further research might be concentrated on two aspects. On one side, it would be interesting to apply our decomposition method to the analysis of inequality of opportunity in other important dimensions of well-being such as income or education. Our method should be fit for these kinds of analysis based on continuous outcomes. Secondly, it would be interesting to analyse equality of opportunity in health using a broader set of circumstances. In this paper, we selected the circumstances judged as most important for health and compatible with our sample size. Further research on a larger data set containing information on parental characteristics might allow us to look at the role of socioeconomic background on the intergenerational transmission of health.

References

Alesina, A., Angeletos, G.M., (2005), Fairness and redistribution, *American Economic Review*, 95: 960-980.

American Diabetes Association, (2010), Executive summary: standards of medical care in diabetes – 2010. *Diabetes Care*, 33: S4-S10.

Atkinson, A., Colburn, W.A., Degruttola, V.G., et al., (2001), Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Biomarkers Definition Working Group, *Clinical Pharmacol and Therapeutics*, 69: 89–95.

Balia, S., Jones A.M., (2008), Mortality, lifestyle and socio-economic status, Journal of *Health Economics* 27: 1-26.

Carrieri, V., Jones, A.M., (2015), The income-health relationship 'beyond the mean': new evidence from biomarkers, *Health, Econometrics and Data Group,* Working paper 15/22, University of York.

Cogneau, D., Naudet, J.D., (2007), Who deserves aid? Equality of opportunity, international aid, and poverty reduction, *World Development*, 35 :104-120.

Contoyannis, P., Jones, A. M., (2004), Socio-economic status, health and lifestyle, *Journal of Health Economics*, 23: 965–995.

Ferreira, F.H.G, Gignoux, J., (2014), The measurement of educational inequality: Achievement and opportunity, *The World Bank Economic Review,* 28 : 210-246.

Ferreira, F.H.G, Peragine, V., (2015), *Equality of opportunity*, in M. Adler and M. Fleurbaey (eds.), Handbook of Well Being and Public Policy, Oxford University Press.

Fleurbaey, M., Schokkaert, E., (2009), Unfair inequalities in health and health care, *Journal of Health Economics,* 28:73–90.

Fleurbaey, M., Schokkaert, E., (2012), Equity in health and health care, *Handbook of Health Economics, Volume 2*: 1003-1092

Gruenewald, T.L., Seeman, T.E., Ryff, C.D., et al., (2006), Combinations of biomarkers predictive of later life mortality, *Proceedings of the National Academy of Sciences,* 103:14158-14163.

Jarvis, M. J., et al. (2008), Assessing smoking status in children, adolescents and adults: cotinine cut-points revisited, *Addiction,* 103: 1553-61.

Jones, A.M., Lopéz Nicolás, A., (2006), Allowing for heterogeneity in the decomposition of measures of inequality in health, *Journal of Economic Inequality*, 4: 347-365.

Jones, A.M., Rice, N., Rosa Dias, P., (2011), Long-term effects of cognitive skills, social adjustment and schooling on health and lifestyle: Evidence from a reform of selective schooling, *Journal of Human Capital*, 5:342-76.

Jones, A.M., Roemer, J., Rosa Dias, P., (2014), Equalising opportunities in health through educational policy, *Social Choice and Welfare,* 43: 521-545.

Jusot, F., Tubeuf, S., Trannoy, A., (2013), Circumstances and effort: how important is their correlation for the measurement of inequality of opportunity in health? *Health Economics*, 22: 1470-1495.

Roemer, J.E., (1998), *Equality of opportunity*, Harvard University Press.

Roemer, J.E., (2002), Equality of opportunity: A progress report, *Social Choice and Welfare,* 19: 455-471.

Rosa Dias, P., Jones, A.M., (2007), Giving equality of opportunity a fair innings, *Health Economics*, 16: 109-112.

Rosa Dias, P., (2009), Inequality of opportunity in health: evidence from a UK cohort study, *Health Economics*, 18: 1057-1074.

Rosa Dias, P., (2010), Modelling opportunity in health under partial observability of circumstances, *Health Economics*, 19: 252 - 64.

Rosero-Bixby, L.S., Dow, W.H., (2012), Predicting mortality with biomarkers: a population-based prospective cohort study for elderly Costa Ricans, *Population Health Metrics*, 10:11.

Sattar, N., Murra, H.M., Welsh, P., et al., (2009), Are markers of inflammation more strongly associated with risk for fatal than for nonfatal vascular events? *Plos Medicine*, 6:1-10.

Sen. A., (2002), Health: perception vs observation, *British Medical Journal*, 13, 324: 860-61.

Trannoy, A., Tubeuf, S., Jusot, F., Devaux, M., (2010) Inequality of opportunities in health in France: a first pass, *Health Economics*,19: 921-938.

World Bank, (2005), *World Development Report 2006: Equity and Development,* Washington.

World Health Organisation (2011), *Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus*, Geneva: World Health Organisation.

Tables

Table 1. Descriptive Statistics – Circumstances

Variables	Percent
Birth Cohorts	
<1959	44.73
1960-1979	39.76
1980+	15.51
Educational level	
below nvq3/gce a	45.09
Nvq 3/gce a level/higher educ.	28.91
nvq4/nvq5/degree or equivalent	26.00
Gender	
Females	48.52
Males	51.48

Table 2 . Distribution of types

Types	Biomarkers			
	Cholesterol	Glycated Haemoglobin	Fibrinogen	Mean Arterial Pressure
1	1,080	1,365	289	1,380
2	794	791	231	834
3	197	196	47	314
4	890	1,300	267	1,322
5	678	709	204	742
6	172	172	38	327
7	403	454	105	462
8	622	618	154	656
9	194	190	43	298
10	523	671	125	684
11	612	632	163	674
12	263	258	58	380
13	374	405	90	411
14	706	701	175	749
15	177	177	30	190
16	493	614	109	621
17	630	650	179	683
18	171	162	29	183
Total	8,979	10,065	2,336	10,910

Table 3. Means of effort variables by type

Type	Cotinine	BMI	Drinking	Medication	Log Income
1	46.53	27.76	13.11	0.71	9.90
2	105.81	26.96	17.29	0.45	9.92
3	106.26	24.30	17.45	0.48	9.62
4	74.60	28.08	25.23	0.68	9.87
5	120.47	27.92	35.23	0.27	10.00
6	99.79	24.37	31.00	0.14	9.86
7	43.95	27.42	14.03	0.61	10.34
8	55.66	26.76	16.70	0.44	10.20
9	60.37	24.24	19.27	0.53	9.88
10	44.60	28.40	25.61	0.61	10.24
11	69.48	27.50	29.51	0.26	10.38
12	52.99	24.60	35.19	0.18	9.91
13	21.37	26.88	17.13	0.58	10.59
14	22.90	25.28	15.79	0.42	10.67
15	15.48	24.11	15.04	0.59	10.52
16	22.19	27.25	24.74	0.57	10.63
17	32.75	26.84	26.20	0.21	10.75
18	25.82	26.17	30.95	0.17	10.46

Table 4. Descriptive Statistics – Biomarkers

Variables	Mean	Std Dev.	Waves[a]	Observations
Cholesterol	5.52	1.09	2,3,4,5,6,7,8,9,10,11	8,979
Glycated haemoglobin	5.55	0.62	2,3,4,5,6,7,8,9,10,11	10,065
Fibrinogen	2.94	0.65	2,3,4,5,8	2,336
Mean Arterial Pressure	77.48	34.69	1,2,3,4,5,6,7,8,9,10,11	10,910

[a] 2002=wave 1; 2012=wave11

Table 5. Regression Results - Pooled OLS and OLS on "the worst-off type"

	Cholesterol		Glycated Haemoglobin		Fibrinogen		Mean Arterial Pressure	
Variable	OLS	Type 1	OLS	Type 1	OLS	Type 1	OLS	Type 1
Cotinine	0.000	0.000	0.002***	0.000	0.001***	0.000	-0.050***	-0.046***
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.002*	*0.007*
BMI	0.047**	0.003	0.026***	0.022***	0.035***	0.023**	0.647***	0.185
	0.002	*0.006*	*0.001*	*0.003*	*0.003*	*0.009*	*0.069*	*0.171*
Drinking	0.001**	0.002	-0.001***	-0.003***	-0.001	-0.006**	-0.005	0.07
	0.000	*0.002*	*0.000*	*0.001*	*0.000*	*0.003*	*0.011*	*0.051*
Medic.	0.067***	-0.129**	0.220***	0.141***	0.234***	0.031	2.332***	-0.344
	0.023	*0.063*	*0.012*	*0.037*	*0.026*	*0.082*	*0.657*	*1.903*
Log Inc.	0.042***	0.027	-0.048***	-0.037*	-0.048***	-0.057	1.110***	0.783
	0.014	*0.04*	*0.008*	*0.022*	*0.015*	*0.046*	*0.411*	*1.139*
Constant	3.774***	5.836***	5.225***	5.400***	2.372***	3.178***	50.714***	70.977***
	0.167	*0.446*	*0.088*	*0.242*	*0.179*	*0.549*	*4.684*	*12.52*
Obs.	8979	1080	10065	1365	2336	289	10910	1380

Standard Errors in Italics; ***, **, * indicate significance at 1%, 5% and 10%, respectively

Table 6. Decomposition Results - Cholesterol

Gini Decomposition										
Variables	**Individual responsibility**	**% of G**	**Direct circumstances**	**% of G**	**Indirect circumstances**	**% of G**	**Group responsibility**	**% of G**	**Total**	**% of G**
Log Inc.	0.000104	0.09	-	-	-0.0096064	-8.62	0.00	0.00		
Cotinine	0.000187	0.17	-	-	0.0000701	0.06	-0.0000344	-0.03		
Drinking	0.000302	0.27	-	-	0.0003461	0.31	-0.0000313	-0.02		
BMI	0.003277	2.94	-	-	-0.0183493	-16.47	0.0016825	1.51		
Medication	0.000444	0.39	-	-	-0.0024236	-2.17	0.0002162	0.19		
Total contribution of variables	**0.004313**	**3.87**	**0.0493546**	**44.30**	**-0.0299631**	**-26.90**	**0.0018409**	**1.65**	**0.025**	**22.93**
Residuals									**0.085**	**77.06**
Gini									**0.111**	
Gini-IOP Decomposition										
Log Inc.	9.35E-05	0.09	-	-	-0.0086467	-8.62	4.51E-06	0.00		
Cotinine	0.000168	0.16	-	-	0.0000759	0.07	-0.0000438	-0.04		
Drinking	0.000271	0.27	-	-	0.000356	0.36	-0.0000725	-0.07		
BMI	0.00295	2.94	-	-	-0.0150968	-15.05	0.0000906	0.09		
Medication	0.0004	0.40	-	-	-0.0016117	-1.61	-0.0003757	-0.37		
Total contribution of variables	**0.003883**	**3.87**	**0.0444373**	**44.31**	**-0.0249233**	**-24.85**	**-0.00039689**	**-0.39**	**0.023**	**22.93**
Residuals									**0.077**	**77.06**
Gini IOP									**0.100**	

Table 7. Decomposition Results - Glycated Haemoglobin

Gini Decomposition										
Variables	**Individual responsibility**	**% of G**	**Direct circumstances**	**% of G**	**Indirect circumstances**	**% of G**	**Group responsibility**	**% of G**	**Total**	**% of G**
Log Inc.	0.0002591	0.52	-	-	-0.015095	-30.15	0.0000000	0.40		
Cotinine	0.0004511	0.90	-	-	-0.000142	-0.28	-0.0000442	-0.09		
Drinking	0.0003398	0.68	-	-	-0.000826	-1.65	0.0000160	0.03		
BMI	0.001753	3.50	-	-	0.0081025	16.19	0.0010819	2.16		
Medication	0.0009651	1.93	-	-	0.0009446	1.89	0.0013899	2.78		
Total contribution of variables	**0.0037681**	**7.53**	**0.0142**	**28.48**	**-0.007016**	**-14.02**	**0.0026429**	**5.28**	**0.013**	**27.27**
Residuals									**0.036**	**72.73**
Gini									**0.050**	
Gini-IOP Decomposition										
Log Inc.	0.000251	0.52	-	-	-0.01458	-30.06	0.0000000	0.30		
Cotinine	0.000437	0.90	-	-	-0.00015	-0.31	-0.0000303	-0.06		
Drinking	0.0003292	0.68	-	-	-0.00084	-1.74	0.0000595	0.12		
BMI	0.0016981	3.50	-	-	0.007995	16.49	0.0009014	1.86		
Medication	0.0009349	1.93	-	-	0.001403	2.89	0.0008586	1.77		
Total contribution of variables	**0.0036502**	**7.53**	**0.0138109**	**28.48**	**-0.00617**	**-12.73**	**0.0019367**	**3.99**	**0.013**	**27.27**
Residuals									**0.035**	**72.73**
Gini IOP									**0.048**	

Table 8. Decomposition Results - Fibrinogen

Gini Decomposition										
Variables	**Individual responsibility**	**% of G**	**Direct circumstances**	**% of G**	**Indirect Circumstances**	**% of G**	**Group responsibility**	**% of G**	**Total**	**% of G**
Log Inc.	0.0008127	0.66	-	-	-0.0144275	-11.74	0.00	0.35		
Cotinine	0.0039524	3.22	-	-	-0.0005659	-0.46	0.0000577	0.05		
Drinking	0.0011498	0.94	-	-	-0.0016866	-1.37	0.0000944	0.08		
BMI	0.0084775	6.90	-	-	-0.0086303	-7.02	0.0013298	1.08		
Medication	0.0021356	1.74	-	-	-0.0012674	-1.03	0.0018965	1.54		
Total contribution of variables	**0.016528**	**13.45**	**0.0367661**	**29.92**	**-0.0265777**	**-21.63**	**0.0038071**	**3.10**	**0.030**	**24.84**
Residuals									**0.092**	**75.16**
Gini									**0.122**	
Gini-IOP Decomposition										
Log Inc.	0.0007468	0.66	-	-	-0.0133315	-11.81	0.00	0.41		
Cotinine	0.0036318	3.22	-	-	-0.0005022	-0.44	0.0000352	0.03		
Drinking	0.0010565	0.94	-	-	-0.002492	-2.21	0.0010289	0.91		
BMI	0.00779	6.90	-	-	-0.0075136	-6.65	0.0008052	0.71		
Medication	0.0019624	1.74	-	-	0.0003451	0.31	0.000233	0.21		
Total contribution of variables	**0.0151875**	**13.45**	**0.0337844**	**29.92**	**-0.0234942**	**-20.81**	**0.0025703**	**2.28**	**0.028**	**24.84**
Residuals									**0.084**	**75.16**
Gini IOP									**0.112**	

Table 9. Decomposition Results - Mean Arterial Pressure

Gini Decomposition										
Variables	**Individual responsibility**	**% of G**	**Direct circumstances**	**% of G**	**Indirect circumstances**	**% of G**	**Group responsibility**	**% of G**	**Total**	**% of G**
Log Inc.	0.0003346	0.16	-	-	0.0026497	1.25	0.000098	0.05		
Cotinine	0.0068108	3.21	-	-	0.0001509	0.07	0.000788	0.37		
Drinking	0.0002791	0.13	-	-	0.000072	0.03	-0.0000090	0.00		
BMI	0.0032326	1.52	-	-	-0.0051554	-2.43	0.0014766	0.70		
Medication	0.0006557	0.31	-	-	-0.0018088	-0.85	0.000413	0.19		
Total contribution of variables	**0.0113128**	**5.33**	**0.0120879**	**5.70**	**-0.0040916**	**-1.93**	**0.0027666**	**1.30**	**0.022**	**10.40**
Residuals									**0.190**	**89.60**
Gini									**0.212**	
Gini-IOP Decomposition										
Log Inc.	0.0003145	0.16	-	-	0.0025041	1.25	0.0000644	0.03		
Cotinine	0.0064221	3.22	-	-	0.0001928	0.10	0.0006931	0.35		
Drinking	0.0002516	0.13	-	-	-0.0000656	-0.03	0.0001239	0.06		
BMI	0.0030554	1.53	-	-	-0.003885	-1.95	0.0003978	0.20		
Medication	0.0006242	0.31	-	-	-0.0012472	-0.62	-0.0000577	-0.03		
Total contribution of variables	**0.0106678**	**5.35**	**0.0114198**	**5.72**	**-0.0025009**	**-1.25**	**0.0012215**	**0.61**	**0.021**	**10.43**
Residuals									**0.179**	**89.57**
Gini IOP									**0.200**	

APPENDIX

Table A.1 Regressions by type - Cholesterol

	Type																	
Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Cotinine	0.000	0.000**	-0.000	0.000	-0.000	0.000	0.000	0.000	-0.000	0.000	0.000	-0.000	-0.000	0.001*	0.001	0.000	0.001*	0.000
BMI	0.003	0.032***	0.039***	0.027***	0.048***	0.093***	0.012	0.027***	0.037**	0.016	0.065***	0.097***	0.006	0.034***	0.028	0.025**	0.065***	0.076***
Drinking	0.002	-0.000	-0.001	0.002**	0.001	0.001	0.002	0.004**	-0.001	0.001	0.002*	0.000	-0.001	0.003	0.001	0.005**	-0.000	0.001
medication	-0.129**	0.031	0.388***	-0.224***	-0.065	0.390*	-0.101	-0.035	0.164	-0.141	0.053	-0.077	-0.188*	0.086	-0.031	-0.265***	-0.038	-0.102
Log Inc.	0.027	0.037	0.096	0.018	-0.016	0.040	-0.050	0.042	0.007	0.010	-0.005	0.075	0.001	0.038	-0.036	-0.077	0.081	0.125
Constant	5.836***	3.905***	2.518***	4.872***	4.451***	1.786*	6.367***	3.957***	3.732***	5.316***	3.779***	1.546***	6.062***	3.759***	4.425***	5.926***	2.846***	1.435
Obs.	1080	794	197	890	678	172	403	622	194	523	612	263	374	706	177	493	630	171

Standard Errors not reported. ***, **, * indicate significance at 1%, 5% and 10%, respectively

Table A.2 Regressions by type - Glycated Haemoglobin

	Type																	
Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Cotinine	0.000	0.000***	0.000*	0.000	0.000***	0.000	0.000**	0.000	0.000	0.000	0.000**	0.001***	0.000	0.000	0.000	0.001***	0.001***	0.002***
BMI	0.022***	0.017***	0.003	0.032***	0.040***	-0.003	0.026***	0.021***	0.011**	0.034***	0.023***	0.014***	0.014**	0.013***	0.003	0.021***	0.008	0.005
Drinking	-0.003***	-0.000	-0.001	-0.002***	0.000	-0.001	-0.003*	-0.001*	-0.001	-0.002**	-0.000	-0.000	-0.003*	-0.001	-0.002	-0.003***	-0.001	0.003**
medication	0.141***	0.091***	0.004	0.239***	0.222***	-0.028	0.127**	-0.016	-0.012	0.244***	0.079*	0.092	0.115**	0.069**	0.012	0.200***	0.202***	0.032
Log Inc.	-0.037*	0.021	0.005	-0.140***	0.025	-0.033	-0.059	-0.027	0.005	-0.062	0.000	-0.021	-0.120***	0.019	-0.015	0.018	-0.008	-0.122*
Constant	5.400***	4.626***	5.084***	6.236***	4.014***	5.724***	5.518***	5.075***	4.938***	5.368***	4.776***	5.061***	6.516***	4.784***	5.311***	4.857***	5.203***	6.297***
Obs.	1365	791	196	1300	709	172	454	618	190	671	632	258	405	701	177	614	650	162

Standard Errors not reported. ***, **, * indicate significance at 1%, 5% and 10%, respectively

Table A.3 Regressions by type - Fibrinogen

	Type																	
Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Cotinine	0.000	0.001**	0.000	0.001***	0.001***	0.001*	0.000	0.000	-0.000	0.001***	0.001***	0.000	0.001*	0.002***	-0.006	0.001	0.001***	-0.000
BMI	0.023***	0.061***	0.045***	0.001	0.024***	0.028	0.018	0.044***	0.050*	0.031**	0.034***	0.015	0.046***	0.044***	0.059*	0.026	0.007	0.045*
Drinking	-0.006**	0.000	-0.008	-0.000	0.001	-0.001	-0.012***	0.000	0.006	0.001	0.001	-0.001	-0.004	0.000	0.002	-0.000	0.002	-0.006*
medication	0.031	0.176**	0.154	0.257***	0.176*	0.464**	0.091	0.153*	0.178	0.109	0.143	-0.038	0.164	0.091	0.127	-0.012	0.343***	0.687
Log Inc.	-0.057	0.021	-0.089	-0.075	0.018	0.044	-0.012	0.031	-0.236**	-0.026	-0.022	0.051	0.013	-0.030	-0.333**	-0.105	-0.008	0.025
Constant	3.178***	0.957*	2.544**	3.729***	1.794***	1.147	2.942***	1.284*	3.896***	2.311***	1.903***	1.713**	1.702	1.864***	5.020**	3.371***	2.426***	1.323
Obs.	289	231	47	267	204	38	105	154	43	125	163	58	90	175	30	109	179	29

Standard Errors not reported. ***, **, * indicate significance at 1%, 5% and 10%, respectively

Table A.4 Regressions by type - Mean Arterial Pressure

	Type																	
Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Cotinine	-0.046***	-0.067***	-0.035**	-0.052***	-0.038***	-0.067***	-0.042***	-0.039***	-0.075***	-0.054***	-0.075***	-0.031*	-0.069***	-0.028**	-0.089**	-0.041***	0.004	-0.013
BMI	0.185	0.288	0.949**	0.390	0.494	1.066**	-0.138	0.265	1.387***	0.820**	0.161	0.902**	0.838**	0.406*	0.357	0.783**	0.719**	1.029*
Drinking	0.070	0.026	-0.046	-0.046	-0.054*	-0.074*	-0.060	0.023	0.022	0.012	0.071*	-0.002	-0.036	0.027	0.021	0.009	0.026	0.055
medication	-0.344	2.540	-2.069	-2.746	0.542	2.614	3.060	2.768	11.769***	-0.922	6.340**	-4.452	4.179	1.169	1.791	-3.342	-1.977	5.130
Log Inc.	0.783	0.510	3.522	0.765	2.954	0.330	-0.175	0.009	1.315	-0.306	1.887	-0.091	-1.978	-2.486	-1.312	2.274	-2.199	2.578
Constant	70.977***	63.957***	10.377	70.071***	36.027*	45.822*	88.614***	67.086***	20.980	65.601***	56.542***	54.324**	80.783***	90.924***	78.895**	40.740*	83.874***	16.137
Obs.	1380	834	314	1322	742	327	462	656	298	684	674	380	411	749	190	621	683	183

Standard Errors not reported. ***, **, * indicate significance at 1%, 5% and 10%, respectively