
Article:
Carrieri, Vincenzo and Jones, Andrew Michael orcid.org/0000-0003-4114-1785 (2018) Inequality of opportunity in health: a decomposition-based approach. Health Economics. pp. 1981-1995. ISSN 1057-9230

https://doi.org/10.1002/hec.3814

Inequality of opportunity in health: a decomposition-based approach

Abstract

This paper presents new decomposition-based approaches to measure inequality of opportunity in health that capture Roemer's distinction between circumstances and effort and are consistent with both compensation and reward principles. Our approach is fully nonparametric in the way that it handles differences in circumstances and provides decompositions of both a rank-dependent relative (the Gini coefficient) and a rank-independent absolute inequality index (the variance). The decompositions distinguish the contribution of effort from the direct and indirect (through effort) contribution of circumstances to the total inequality. Our approach is illustrated by an empirical application which uses objectively measured biomarkers 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 and in a combined ill-health index while effort plays a substantial role in explaining inequality in fibrinogen only.

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

JEL codes: C1, C5, D63, I14.

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 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 (EOp). Following Roemer's framework (Roemer, 1998, 2002; Roemer and Trannoy, 2016), this literature separates the factors associated with an outcome of interest into two components: 'circumstances', which are not under individual responsibility, and 'efforts', for which to some extent they are held responsible.

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, et al., 2011; Jones, et al., 2014) to policies related to the allocation of the international aid to countries for the reduction of poverty (Cogneau and Naudet, 2007). The theoretical relevance of equality of opportunity in health has been advocated by many authors (e.g., Sen, 2002; Rosa Dias and Jones, 2007; Fleurbaey and Schokkaert, 2009, 2012) and the relevance of equality of opportunity has been placed at the top of the 'inequality of what' debate by relevant institutions (e.g. World Bank, 2005). A growing literature has addressed the measurement of inequality of opportunity (IOp) in health (e.g., Rosa Dias, 2009; Rosa Dias, 2010; Trannoy et al., 2010; Jusot, et al., 2013; Li Donni, et al., 2014; García-Gómez et al., 2015; Kim, 2016).

In this paper we propose decomposition-based approaches to measure inequality in objective health that capture Roemer's distinction between circumstances and effort. We fully condition on circumstances by splitting our sample according to 'types', who share the same circumstances, and then estimating separate regressions of health outcomes on effort for each sub-sample. This approach allows the model to be fully nonparametric in the way that it handles the circumstances. Using linear regression within the sub-samples generates a heterogeneous set of regression coefficients that we use in a regression-based decomposition of total inequality in the health outcomes. A valuable feature of this method is that it is able to consider simultaneously the two main views existing in the literature on inequality of opportunity: the *compensation principle* and the *reward principle*. Interestingly, this does not require additive separability but only linearity in effort of our health production function. The fact that our approach allows the possibility of interaction between circumstances and effort (through heterogeneous slopes) is also relevant for the assessment of *direct unfairness* and the *fairness gap* in the spirit of Fleurbaey and Schokkaert (2009, 2012).

To retrieve the relative contribution of circumstances and effort to total inequality, we first exploit a decomposition of the Gini coefficient with heterogeneous responses proposed by Jones and Lopez-Nicolas (2006). Moreover, we also propose an alternative derivation of our decomposition terms based on the variance decomposition formula, proposed by Shorrocks (1982) and adapted to the measurement of inequality of opportunity by Jusot, et al. (2013) [1]. Our decomposition methods identify four normatively-relevant decomposition terms: a direct and an indirect (through effort) contribution of circumstances to the total inequality, the contributions of effort to total inequality, and the contribution of residual variation within types.

We illustrate our methods with an empirical application and a second contribution is the use of biomarkers as outcome variables and as proxies of relevant effort variables. [2] As health outcomes, we consider four biomarkers that are associated with some of the most prevalent diseases in all Western countries: cholesterol, glycated haemoglobin, fibrinogen and a combined ill-health index (the first component of a principal component analysis

[1] Rank dependent inequality indices, such as the Gini coefficient and the concentration index, have been the workhorse in the health economics literature on socioeconomic inequalities in health (e.g., Wagstaff et al., 2003; Fleurbaey and Schokkaert, 2009). The Gini and its variants are also used in analyses of inequality of opportunity (e.g., Aaberge et al., 2011; Abatemarco, 2015; Rosa Dias, 2009; Trannoy et al., 2010; Turk and Östh, 2017). The variance has been advocated as an appropriate measure of inequality in health outcomes by Fleurbaey and Schokkaert (2009) and applied in Jusot, et al. (2013), García-Gómez et al. (2015) and Kim (2016). Other indices that have been used to measure inequality of opportunity for monetary or health outcomes include the Theil index (e.g., Bourguignon et al., 2007), Atkinson index (e.g., Li Donni et al., 2014), and mean logarithmic deviation (e.g., Turk and Östh, 2017).

[2] Biomarkers are characteristics that are 'objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention' (Atkinson et al., 2001). 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).

on the three biomarkers). As effort variables, we use saliva cotinine, a major metabolite of nicotine, to objectively quantify individual smoking, along with detailed self-reported data on intensity and frequency of drinking behaviour and the portions of fruits and vegetables consumed as a proxy for a healthy diet. 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 differences in reporting behaviour across individuals. Indeed, previous empirical investigations show systematic variation in reporting across socio-economic groups which may bias the estimates of the inequality of opportunity in health in a significant way (e.g., Sen, 2002).

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

2. Methods

2.1 A Normative Framework

To model inequality of opportunity in health we adopt the framework of Roemer (2002). Roemer partitions 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 judged to be beyond individual responsibility, are regarded as a source of unfair differences in outcomes.

A general health production function can be defined along the lines of Roemer (2002) as $\mathrm{H}(\mathrm{C}, \mathrm{E}(\mathrm{C}))$ where C denotes individual circumstances and E denotes effort, which is itself a function of circumstances. To reflect the fact that observed realisations of health outcomes are inherently random and that the equality of opportunity ethic can be expressed in terms of factors associated with the distribution of health, this is written in terms of the distribution function of the realised individual outcomes conditional on observed circumstances and effort:

$$H_i \sim H\big(C_i, E_i(C_i)\big) \tag{1}$$

where H_i denotes the health outcome for the *i*th individual, and C_i, E_i their circumstances and effort, respectively.

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.[3] 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.[4] This implies that, in addition to assuming a partitioning between C and E, our model assumes that effort is a function of circumstances. It also assumes that circumstances are pre-determined and should not be a function of effort.

The key to our method is that we condition on circumstances by splitting our sample according to type, τ, and then estimate separate regressions of health outcomes on effort for each sub-sample. Conditioning on circumstances gives the following distribution functions, within each type:

$$H_i \sim H\big(C_i, E_i(C_i)\big) = F_\tau(E_i) \; for \; all \; \tau = 1, \dots, T \tag{2}$$

[3] We follow the traditional approach in the IOp literature based on the *a priori* selection of the types according to the circumstances variables. Recently, Li Donni, et al. (2015) proposed a more data-oriented approach for the definition of social types based on the estimation of latent class models.

[4] This ethical stance is open to dispute. For example Jusot et al. (2013) propose empirical methods to compare the Roemer view with two more libertarian perspectives: first, that individual efforts should be fully respected whatever the influence of circumstances on those efforts; second, that regards the efforts of earlier generations, especially parents, to improve outcomes for their offspring as legitimate. So circumstances should only include factors that do not shape effort.

Now assume that, within each type, the realised health outcome generated by the distributions specified in (2) is a linear function of efforts and a random error term:

$$H_i = \alpha_\tau + \beta_\tau E_i + u_i^\tau \quad (3)$$

Equation (3) 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).[5] It is important to note that equation (3) does not require additive separability of circumstances and effort and allows interactions between them (through the heterogeneous slopes that vary with circumstances β_τ).

The u_i^τ are the type-specific error terms, capturing the unexplained variation, within-types, due to the contribution of unobserved factors that are not captured by the observed C and E variables. Note that the mean of the unobservables within-types will be subsumed into the intercept terms, α_τ and any correlation with effort within type will be subsumed in β_τ.[6] Also the variance and other higher moments of the error term are allowed to vary across observed types.

So, to recap, our approach relies on three normative assumptions:

(i) the partitioning of circumstances and effort;

(ii) that effort is a function of circumstances and not *vice versa* and;

(iii) that, conditional on circumstances (type), there is a linear relationship between effort and outcomes.

Assumptions (i) and (ii) are standard for the Roemer model (Roemer, 1998, 2002; Roemer and Trannoy, 2016) and widely adopted in empirical applications. Assumption (i) requires a complete and non-overlapping partition of observed factors between circumstances and effort, sometimes referred to as a dichotomic outcome function (Lefranc et al., 2009; Abatemarco, 2015).[7] Here we apply assumption (i) to factors associated with the distribution of effort, allowing the actual realisations of health to be random (due to the addition of error terms u_i^τ that are type-specific). As discussed above, systematic differences in the mean of the error term across observed types are absorbed into α_τ and in the empirical application we investigate whether there is systematic variation in the higher moments – such as the variance – that is correlated with observed circumstances or effort[8].

Assumption (ii) allows effort to be shaped by circumstances but not *vice versa*. This can be interpreted in terms of the direction of causality and also in terms of the control view of inequality of opportunity: that circumstances are factors which are not the individual's responsibility while, to some extent, efforts are. In practice, the partitioning of variables between C and E should respect this assumption.

Assumption (iii) relaxes the common assumption that outcomes (or the latent variables in models for binary outcomes) are linear in both effort and circumstances that has been used widely in applied work on measurement of inequality of opportunity (e.g., Bourguignon et al., 2007; Trannoy et al., 2010; Jusot et al., 2013; Garcia-Gomez et al., 2015). Our approach is therefore semiparametric in the spirit of Li Donni et al. (2015) although our statistical approach differs from theirs.

[5] In order to keep the notation simple, equation (3) 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).

[6] Garcia-Gómez et al. (2015, p.1354) share this view on the omitted variable issue: "However, from the point of view of inequity measurement, this is a not a problem. On the contrary, it is better that the estimated association between C and the dependent variable of interest reflects to some extent the effect of equity-relevant but unobservable variables". They express doubt about omitted effort variables but recall that under the Roemer perspective the distribution of effort that is correlated with circumstances is regarded as a circumstance itself.

[7] In contrast, some studies in this literature take a non-dichotomic approach and distinguish a further set of factors, such as demographic characteristics, that are treated as controls but are not given a normative significance as either legitimate or illegitimate sources of inequality (eg., Jusot et al., 2013; Roemer and Trannoy, 2016). Others have emphasised the role of random factors, or 'luck' (e.g., Lefranc et al., 2009). Lefranc et al. (2009) argue that luck may be a legitimate source if it is 'even-handed' in the sense of being uncorrelated with circumstances. The Lefranc et al. (2009) approach justifies the use of stochastic dominance criterion to test for IOp, especially when it is combined with Roemer's approach of defining tranches of effort in terms of an individual's relative rank in the distribution within types.

[8] In our empirical application we find very little correlation between higher moments of the type-specific residuals and both observed circumstances and effort variables.

2.2 Decomposition by Factor Components

To retrieve the contribution of circumstances and efforts to total inequality we first exploit the method proposed by Jones and Lopez-Nicolas (2006) who show how regression-based decomposition methods for the decomposition of health inequality, for example as measured by the Gini index, can be extended to incorporate heterogeneity in the responses of health to the explanatory variables (as in equation (3)).[9] 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.[10] Then, we show that our decomposition method applies to the variance decomposition of Shorrocks (1982).

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

$$G = \frac{2}{\overline{H}}\ Cov(H_i, R_i) \tag{4}$$

Or:

$$G = \left(\frac{2}{N\overline{H}}\right)\sum_i (H_i - \overline{H})\left(R_i - \frac{1}{2}\right) \tag{5}$$

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').

To provide a benchmark for our decomposition analysis, first define the *effort* of individual *i* as the product of the effort variable and the associated slope parameter:

$$B_i = \beta_\tau E_i \quad (i \in \tau) \tag{6}$$

Then as a benchmark use the weighted averages across types, where $\pi_\tau$denotes the share of each type:

$$\overline{H} = \sum_\tau \pi_\tau \overline{H}_\tau\ ,\ \ \overline{B} = \sum_\tau \pi_\tau \overline{B}_\tau\ ,\bar{\alpha} = \sum_\tau \pi_\tau \alpha_\tau\ ,\bar{\beta} = \sum_\tau \pi_\tau \beta_\tau \tag{7}$$

Now, consider that:

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

where, given our linear specification of Equation (3):

$$(H_i - \overline{H_\tau}) = (B_i - \overline{B}_\tau) + u_i^\tau \tag{8}$$

and:

$$(\overline{H_\tau} - \overline{H}) = (\alpha_\tau - \bar{\alpha}) + (\overline{B_\tau} - \overline{B}) \tag{9}$$

Following Jones and Lopez-Nicolas (2006), we can substitute (3) into (5). Then, by substituting (8) and (9) and changing the order of summations, the decomposition of the Gini index can be expressed as follows:

---

[9] This approach is based on decomposing inequality indices for an outcome that can be expressed as a (weighted) sum of a set of factor components (e.g., Rao, 1969; Shorrocks, 1982). This was extended to the decomposition of outcomes that can be expressed in terms of a linear regression model in Wagstaff et al,. (2003) and applied to health concentration indices. Jones and Lopez (2006) extend their approach to allow for heterogeneous regression coefficient. The Wagstaff et al. (2003) regression-based decomposition approach for concentration indices has been criticised on various grounds. First, rank ignorability, that it treats the ranks, R, as fixed and does not allow for the influence of covariates on the ranking variable (Erreygers and Kessels, 2013; Heckley et al., 2016). Second, weighting function ignorability which, for the Gini coefficient, implies treating the mean of the outcome as fixed and not allowing for the influence of covariates on the mean (van Ourti et al., 2009; Heckley et al., 2016). Third, Heckley et al. (2016) are critical that the regression-based approach does not explicitly define counterfactual values of the covariates. They propose a method that uses recentred influence functions (RIF) to estimate the partial effect of covariates on a general class of concentration indices but that does not provide a complete percentagewise decomposition of the indices.

[10] It is worth noting that our method consists of a by-factors decomposition of the Gini Index. A by-group decomposition of the Gini Index for the measurement of inequality of opportunity in income has been recently adopted by Abatemarco (2015). While Li Donni et al. (2014) work with a by-group decomposition of the Atkinson index.

$$G = \left(\frac{2}{N\bar{H}}\right)\sum_{\tau}\sum_{i\in\tau}(\alpha_\tau - \bar{\alpha})(R_i - \tfrac{1}{2}) + \left(\tfrac{2}{N\bar{H}}\right)\sum_\tau \sum_{i\in\tau}(\bar{B}_\tau - \bar{B})(R_i - \tfrac{1}{2}) + \left(\tfrac{2}{N\bar{H}}\right)\sum_\tau \sum_{i\in\tau}(B_i - \bar{B}_\tau)(R_i - \tfrac{1}{2}) + \left(\frac{2}{N\bar{H}}\right)\sum_{\tau}\sum_{i\in\tau} u_i^\tau (R_i - \tfrac{1}{2}) \quad (10)$$

The first term in equation (10) is the contribution of the variation of the intercepts of the OLS regression across types (centred at the pooled mean)[11]. In normative terms, this measures the *direct contribution of circumstances* to the overall inequality. The second term relates to variation in the average level of effort within each type around the pooled mean of effort, it therefore measures the *indirect contribution of circumstances* to overall inequality, through differences in the association between efforts and outcomes across the types[12]. The third term measures the contribution of within-type variation in effort to overall inequality. In normative terms, this represents the contribution of *effort* to the overall inequality. The final term is the contribution of the within-type error term and it measures the contribution of *residual factors* to overall inequality. [13]

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 (Roemer, 1998, 2002). The resulting inequality index – which we call an Inequality of Opportunity Gini - is thus expressed in terms of inequality relative the most disadvantaged type. The resulting decomposition terms follow the same logic of those in equation (10) but they are expressed with reference to the situation of the worst-off type. The detailed derivation of this decomposition is provided in the Appendix.

A potential limitation of the regression-based decomposition methods illustrated so far is that they both rely on the *rank ignorability* and *weighting function ignorability* assumptions which have been criticised for being rather restrictive in the analysis of health inequalities (Heckley et al., 2016). In the Appendix we present an alternative decomposition using the absolute Gini and we show that as long as the percentagewise decomposition is the main focus of attention – as it is here - *weighting function ignorability* is not a concern.[14] We do not therefore present separate results for the decomposition of the absolute Gini.

Some recent contributions to the literature on inequality of opportunity in health have favoured the variance as an absolute (rank-independent) measure of health inequality (see e.g., Fleurbaey and Schokkaert, 2009, 2012; Jusot, et al., 2013; García-Gómez et al., 2015; Kim, 2016). To reflect this, we follow Jusot et al. (2013) and propose a derivation of our method based on the variance decomposition of Shorrocks (1982) which is illustrated in detail in the Appendix. The resulting decomposition terms follow the same logic described for the

[11] Note that the overall Gini coefficient can be expressed as a scaled covariance and that, in the decomposition, each of the contributions is an expression that is analogous to a covariance: given by the sample mean of the product of deviations of a quantity around a mean value and the deviation of the relative rank variable around its mean.

[12] One point to note is that the separation of direct and indirect components may depend on the scaling of the effort variables, E. The intercept terms in the direct contribution correspond to the reference point where $E = 0$ so shifting the location of E would affect the relative size of direct and indirect contributions. So long as effort is measured on a ratio scale the relative sizes do not change as a result of a rescaling. In our empirical application we set the reference level of effort, where $E = 0$, to correspond to the highest level of effort that can be achieved. This follows the spirit of Fleurbaey and Schokkaert (2009) and applications in García-Gómez et al. (2015) and Kim (2016) that use the best levels of C and E as reference points reflecting the implicit norm of vertical equity that this implies.

[13] To check for any higher-order residual correlation we conduct an auxiliary regression where we regress $\hat{u}_i^\tau * R_i$ on the observed types and the effort variables.

[14] The invariance of the percentage decomposition also applies to the numerous variants of rank dependent indices that have been proposed in the recent literature such as the Erreygers, Wagstaff, Attainment Relative and Shortfall Relative indices (Erreygers, 2009; Heckley et al., 2016). In addition, a nice feature of our empirical application arises from the cardinality of our measures of health, i.e. the biomarkers. For example, Van Doorslaer and Jones (2003) note that in such a situation, the percentage factor contributions in the Gini decomposition remain unchanged under different cardinal transformations of the outcome variable and the decomposition is invariant to linear transformations of the outcome. This is a desirable property of a factor decomposition of any rank-dependent inequality index (see, van Doorslaer and Jones, 2003; Erreygers, 2009).

Gini decomposition and have the same normative interpretation with direct and indirect contributions of circumstances, effort and the residual component.[15]

Our approach decomposes explained inequality into terms that relate to both compensation and reward principles[16]. The sum of all sources of explained inequality deserving compensation (*direct and indirect circumstances*) corresponds to the inequality which remains when legitimate sources of inequality (the *effort term*) are deducted from total explained inequality under the reward principle. The compatibility of our approach with both the compensation and the reward principles relies on the assumption of linearity in circumstances and effort of our health outcome function (see equation 3). It has been demonstrated that in general inequality of opportunity measures can be either fully consistent with the reward or with the compensation principle but not necessarily both (see for instance Fleurbaey, 2008). Our approach represents an appealing compromise to this trade-off, because despite our linearity assumption, we allow for interaction effects between circumstances and effort, through the heterogeneous slopes on the effort factors across types.

2.3 Empirical Implementation

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, et al., 2013; García-Gomez et al., 2015) and the normative literature on the measurement of health equity (see the discussion in Rosa Dias and Jones (2007) for more details). Thus, we treat as circumstances the cohort of birth, gender, educational level, and neighbourhood (more *vs* less deprived areas) based on the index of multiple deprivation scores.[17] In the case of education, we assume that the level of secondary schooling achieved by age 18 is beyond their individual responsibility and therefore constitutes a circumstance. This is an assumption shared by other papers (e.g. Rosa Dias, 2010). Moreover, we also assume that the residential status in more *vs* less deprived areas is beyond the individual responsibility. This point has been nicely discussed by Burchardt and Le Grand (2002) who place residential area among circumstances virtually modifiable by individuals but "with very high social, psychological and financial costs". Given the relevance of neighbourhood conditions for health status (e.g., Bilger and Carrieri, 2013) we opted to include it among circumstances factors in our analysis. Moreover, its inclusion among circumstances is useful in our application to take into account also the social background of individuals.

Following this strand of literature, the choice of effort variables is guided by work on the relationship between health and lifestyles (e.g., Contoyannis and Jones, 2004; Balia and Jones, 2008; García-Gómez et al., 2015). Lifestyles are determined by individual decisions to invest in health capital, and, therefore, they are, at least partly, within individual control. Thus, we treat cigarette smoking (saliva cotinine), alcohol frequency and intensity of consumption and dietary choices (consumption of fruit and vegetables) as effort factors.

3. Data

We use ten waves (2003-2012) of the Health Survey for England (HSE).[18] In the HSE, the interview includes a set of core questions, asked each year, on general health and psycho-social indicators, smoking, alcohol,

[15] To check for any higher-order residual correlation we conduct an auxiliary regression where we regress $\hat{u}_i^{\tau} * H_i$ on the observed types and the effort variables.

[16] A careful discussion around the compensation and rewards principles in the IOp framework can be found in Fleurbaey, 2008; Aaberge, et al., 2011 and Li Donni, et al., 2014. These two approaches have a clear parallel with the fairness gap and the direct unfairness approaches proposed by Fleurbaey and Schokkaert (2009) for the specific case of health inequalities. Comprehensive discussions on the different approaches to measure IOp can be found in Checchi and Peragine (2005; 2010) while a careful review of the different approaches and measurement issues of IOp can be found in Ferreira and Peragine (2015).

[17] García-Gómez et al. (2015, p.1364) argue that age and gender "should be considered explicitly in any analysis of inequality of opportunity".

[18] 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 average agreement rate is quite high (close to 60%) and does not show a systematic pattern across the social types defined in our analysis. This mitigates potential sample-selection concerns.

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.

We use the valid (i.e. blood sample properly collected and successfully processed) biomarker measurements in each wave. Thus, we can use 11,096 non-missing observations for the analysis of cholesterol over the period 2003-2012 and 12,516 for the analysis of glycated haemoglobin over the period 2003-2012. The sample size for fibrinogen is much smaller yielding 2,795 observations from 2003 to 2006 and in 2009. Similarly, the sample size for our combined ill-health index is 2,724 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 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 and some descriptive statistics. A glossary of all biomarkers used in our empirical investigation along with their clinical cut-points is reported in the Appendix.

Circumstances (C)
We use four variables to define circumstances: cohort of birth, gender, individual education and area of residence. 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 used to split the sample in three categories according to the level of secondary schooling attained: completing only compulsory secondary schooling (qualification below nvq3/GCE A level); completed secondary schooling (nvq 3/GCE A level); continued to further/higher education (nvq4/nvq5/degree or equivalent). As a short-hand we refer to these as low (LE), moderate (ME) and high (HE) levels of education in the tables of results. Area of residence refers to the deprivation of the area of residence based on the scores of the index of multiple deprivation (IMD).[19] We split the sample in two categories: higher-deprived (HD as a short-hand in the table of results) and lower-deprived (LD) based on whether individuals live in an area belonging to the top two quintiles of the IMD score.

A summary of these variables is presented in Table 1. This shows that around 51% of the sample were born before 1959, around 32% were born between 1960 and 1979 and around 16% were born in 1980 or later. The figures are indicative of the ageing population common to many European countries. Table 1 also shows that around 54% of our sample completed only compulsory secondary education, while around the 45% of the sample are men and the share of individuals living in more deprived areas is around 36%.

On the basis of the combination of the circumstances discussed above, we can define 36 types. These are described in Table A.1 in the Appendix along with their distribution in our sample. Type 1 is the type for which we might expect *a priori* to have the greatest disadvantage in terms of health outcomes: born before 1959, with lower education, female and living in a more deprived area. Conversely, type 36 is the type for which we might expect the greatest advantage: born in 1980 or later, highly educated, male living in a less deprived area. In practice we use an *ex post* approach, following Roemer (2002), to define the most disadvantaged type for each health outcome based on the empirical distribution function of biomarkers across types.

As Table A.1 shows, on average, we have a reasonable sample size within each type for cholesterol, and glycated haemoglobin and, importantly, we have a relatively large sample size for most types. Both aspects are relevant for our empirical analysis which is based on estimating separate regression of health outcomes on effort for each sub-sample. For fibrinogen and the ill health index, we have a much smaller average sample size (2,795 and 2,724

[19] The index of multiple deprivation is a measure of relative deprivation for small areas (Lower Super Output Areas (LSOAs) of about 1,500 inhabitants) and it is a combined measure of deprivation based on a total of 37 separate indicators that have been grouped into seven domains (income, employment, health deprivation and disability, education, barriers to housing and services, crime and living environment) each of which reflects a different aspect of deprivation experienced by a given area.

observations respectively) and some sub-samples have few observations. For these reasons, the results related to fibrinogen and the ill health index should be interpreted with more caution.

Table 1. Descriptive Statistics – Circumstances

Variables	Percent
Birth Cohorts	
<1959	51.22
1960-1979	32.61
1980+	16.17
Educational Level	
Low (LE)	54.13
Moderate (ME)	25.30
High (HE)	20.57
Gender	
Females	54.17
Males	45.83
Deprivation score	
High (HD)	36.08
Low (HD)	63.92

Efforts (E)

As effort variables, we consider health-related behaviours: smoking, diet, and drinking. As a proxy of smoking, we use saliva cotinine (more details are provided in the Appendix). As a second effort factor, we use the portions of fruits and vegetables consumed in the day before the interview as a proxy for a healthy diet. To be consistent with the other effort variables we re-scale fruit and vegetable consumption so that zero corresponds to a maximum effort of 10 portions per day and the variable is labelled as "bad diet" in the tables of results.[20] As a proxy of drinking behaviour, we use self-reported information on 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.

In Table A.2 in the Appendix, 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. Similarly, Diet changes quite substantially across types. On the contrary, drinking follows a less sharp pattern across types.

Health outcomes (H)

Table 2 shows the descriptive statistics of the biomarkers. 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 cut-point of 5 while fibrinogen average scores are a little lower than the normal cut-point of 3. Moreover, Table 2 shows higher dispersion around cholesterol and ill health index scores, while other biomarkers values are less dispersed around the mean. Poorer health outcomes tend to be more concentrated among disadvantaged types (see Table A.3 in the Appendix). In particular, average biomarker levels of type 1 are substantially above the clinical threshold in the case of cholesterol and fibrinogen. Conversely, biomarker levels of type 36 all fall within the normal ranges. In between, we observe a higher concentration of bad health outcomes especially among types made up of older individuals.

[20] The reference point of 10 portions reflects evidence from a recent meta-analysis of the benefits of fruit and vegetables (Aune et al., 2017).

Table 2. Descriptive Statistics – Biomarkers

Variables	Mean	Std Dev.	Waves[a]	Observations
Cholesterol	5.52	1.10	1,2,3,4,5,6,7,8,9,10	11096
Glycated haemoglobin	5.56	0.63	1,2,3,4,5,6,7,8,9,10	12516
Fibrinogen	2.93	0.65	1,2,3,4,7	2795
Ill Health Index*	3.81	1.19	1,2,3,4,7	2724

[a] 2003=wave 1; 2012=wave10 *First component of a PCA on Cholesterol, Glycated Haemoglobin and Fibrinogen.

On the basis of the empirical distribution functions we can identify the worst-off type for each biomarker and we use this type as a benchmark (in equation 13) for the Inequality of Opportunity Gini decomposition. These are types 25, 8, 13 and 31 for cholesterol, glycated haemoglobin, fibrinogen and the ill health index, respectively. All these types are composed by older individuals (born before 1959) and by people living in more deprived areas (with the only exception of type 8). This suggests that age remains the most relevant risk factor for health and also that deprivation of the area plays an important role in shaping health outcomes.

4. Empirical Application

In Tables A.4-A.7 in the Appendix, we report the complete set of regressions for all 36 types and for all biomarkers analysed. The main results of these regressions are that effort variables generally have significant slope coefficients but display a large degree of heterogeneity across biomarkers and across types. Importantly, we found a large heterogeneity across types with respect to the constant (direct) terms and this anticipates that circumstances play a large direct role in influencing health outcomes. With respect to the heterogeneity across slopes, we found that the worse-off types generally exhibit significantly higher slope coefficients, especially for the effect of unhealthy diet.

The results of our decomposition analysis are reported in Tables 3-6 for cholesterol, glycated haemoglobin, fibrinogen and the ill health index, respectively. In each table, we report the decomposition of the Gini index (in the top panel) and of the Variance (bottom panel) into the four contributions. Results of the IOp-Gini decomposition are not reported since they are very similar to the results of the standard Gini decomposition. All terms are expressed in units and as a percentage of the explained inequality indices. As the sign and the magnitude of the contributions are very similar across the three inequality measures, we also illustrate in Figure 1 the contribution of the decomposition terms for the Gini index decomposition only. A number of robustness checks considering additional circumstances variables (ethnicity and parental smoking status) and effort variables (use of prescribed medications) leave our results substantially unchanged and are discussed in detail in the Section IV in the Appendix.

[Insert Tables 3-6 around here]

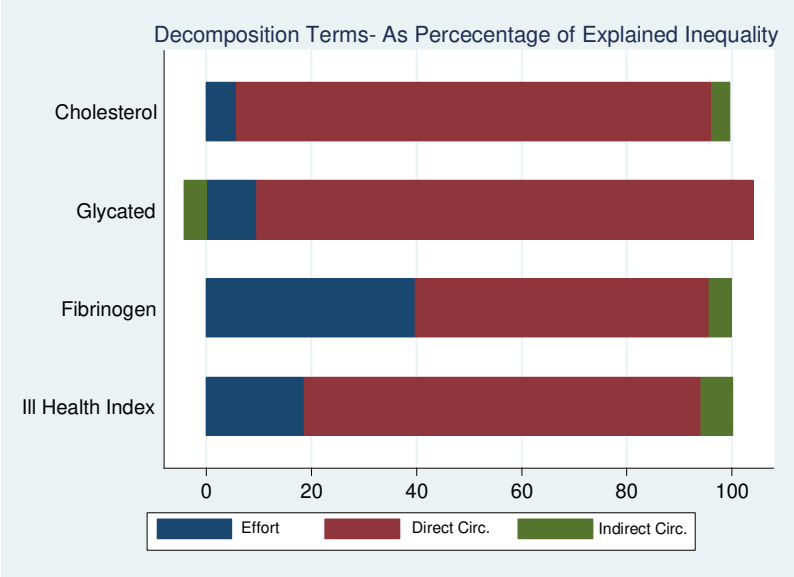
Tables 3-6 and Figure 1 show that the largest contribution to explained inequality in the outcomes is attributable to a direct effect of circumstances for all of the biomarkers. Indeed, the direct contribution of circumstances ranges from around 56% of explained inequality in fibrinogen to around 95% of explained inequality in glycated haemoglobin.

The second contribution to inequality is attributed to effort. For cholesterol and glycated haemoglobin its contribution is however only marginal (i.e. around 6% and 9% of the explained inequality, respectively) while for the other biomarkers, the contribution of effort terms is much more important and it reaches around 18% of the explained inequality in Ill-Health Index and around 40% of the explained inequality in the case of fibrinogen.

The contributions of the indirect circumstance terms is less important and ranges from around 3.5% for cholesterol to around 6% for Ill-Health index. Its contribution is positive for all biomarkers with the exception

of glycated haemoglobin. This implies that the interaction between circumstances and effort, through the slope coefficients, generally increases the overall level of inequality. As discussed before, this reflects the fact that the types that, on average, have higher biomarker scores (i.e. worse health) often have higher slope coefficients on the measures of effort. In the case of glycated haemoglobin, the negative terms indicated that types in poorer health have lower slope coefficients.

Figure 1. Decomposition Results- Gini Index



The patterns described above are essentially common to all biomarkers. Only fibrinogen is an exception with respect to the role of effort terms which is significantly larger than for the other biomarkers, and glycated haemoglobin with respect to the indirect circumstance terms which is negative (while it is positive for all the other biomarkers). Despite these minor exceptions, the ranking of the contributions is the same for all biomarkers analysed and this indicates that there is a general pattern of the contributions to inequality which is common to all of the health outcomes.

For overall inequality, we find that their levels are heterogeneous across biomarkers. Overall inequality is generally low for glycated haemoglobin while it is higher for cholesterol and fibrinogen and significantly higher for the ill-health index . This is consistent with the fact the latter reflects variations in a broader range general health conditions, i.e. it takes into account all biomarkers considered. A direct comparison between the estimated Gini measures and the variance is not appropriate but also in the case of variance, higher dispersion is found for the ill-health index and cholesterol (in order of magnitude) while lower dispersion is found for fibrinogen and glycated haemoglobin. Importantly, both the magnitude, the sign and ranking of the contributions are very similar under all the indices used and for all biomarkers analysed.

Furthermore, we observe a very large contribution of the unexplained component for all of the biomarkers. This is largest in the case of fibrinogen, amounting to around 80% of the overall inequality. For the other biomarkers, the contribution is a little lower, ranging from around 71% to 79%. As expected, this demonstrates that observed circumstances and efforts offer only a partial explanation of the overall observed variation in the realised outcomes for the biomarkers. This is a common feature of studies dealing with the measurement and decomposition of inequality in health. However we find very little correlation between higher moments of the type-specific residuals and both observed circumstances and effort variables[21]. This supports the idea that unexplained component can be mostly regarded as a random noise in our empirical application.

[21] F-tests of the joint significance of observed circumstances and effort variables are 0.71 and 0.16 for Gini, and variance decomposition residuals, respectively, in the cholesterol regression; 1.09 and 0.58 in glycated haemoglobin regression; 0.47 and 0.18 in fibrinogen regression; 0.39 and 0.24 in ill-health index regression.

Lastly, Tables 3-6 show the detailed contribution of each effort variable to the overall inequality. We find that, in terms of *effort terms*, all the effort variables plays a roughly equal role in the overall inequality across the majority of biomarkers analysed. Drinking is a little more important for glycated haemoglobin while cotinine dominates for fibrinogen and the ill-health Index. For what concerns the detailed decomposition of the *indirect circumstances* we find that unhealthy diet and drinking represent the most important effort variables. The contribution of unhealthy diet is generally positive for all biomarkers. This implies that the slope coefficient for unhealthy diet is larger for types that have higher rankings for the biomarkers and hence worse health (typically the worst-off types). In the case of glycated haemoglobin only, the contribution is negative, meaning that slope coefficient for this variable is larger for types with a less diabetes risk. The contribution of drinking behaviour is positive for cholesterol and glycated haemoglobin and it is negative for the other biomarkers. The contribution of smoking is more negligible for all biomarkers.

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 associated with 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 that measures inequality relative to the most disadvantaged type. Moreover, we have also shown that our decomposition method applies to the variance decomposition of Shorrocks (1982).

We illustrate our method with an application to the analysis of inequality of opportunity in three biomarkers that are associated with some of the most prevalent non-communicable diseases: cholesterol, glycated haemoglobin, fibrinogen along with a general ill-health index built on the combination of the all three biomarkers. Moreover, we use a biomarker to measure smoking. The use of biomarkers is new in the analysis of health equity and it is useful to have measures of inequality of opportunity that are not biased by reporting heterogeneity.

Using ten waves of the Health Survey for England (2003-2012), we find that the target of equality of opportunity in health is still far from being reached in England. Our investigation shows that circumstances are still a key source of health inequalities for all health outcomes analysed explaining from 56% to 95% of the total inequality. In some cases, i.e. for diabetes risk, we find, in addition, a significant interaction between circumstances and efforts. In the case of glycated haemoglobin, this interaction has the effect of dampening overall inequality. This is an aspect that should be carefully considered in inequality of opportunity analyses which often rely on the hypothesis of the separability of circumstances and effort. Moreover, this result has potential policy implications as it suggests that the possibility of decreasing inequalities through higher individual efforts may be limited in the case of more disadvantaged individuals and for specific diseases such as diabetes.

At the same time, we find that individuals are still empowered to reduce the risks for some specific diseases. Individual effort, and in particular smoking and drinking behaviours are found to be very important for the risk of inflammatory diseases, associated with higher fibrinogen levels. For the latter, we find a contribution of effort which is almost equal to the direct role of circumstances. Similarly, people in worse circumstances are empowered to reduce the risk of some diseases through healthier eating behaviour due to a steeper association between unhealthy diet and health among people in worse circumstances. All in all, our results suggest that health policy interventions designed to encourage the adoption of healthy lifestyles may have limited effectiveness or be effective only for the prevention of specific diseases, while a wider strategy aimed at equalising opportunities would be needed to substantially reduce inequalities in health.

Our decomposition method offers the possibility of extensions and further applications. First, 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. Second, further research might strongly benefit of a combination of different

kinds of data. Because of some data limitations, our empirical illustration is not able to address the role of parental background, other than parental smoking, on the intergenerational transmission of health which is generally analysed in inequality of opportunity studies. Similarly, there is a wide literature (see Goodman et al, 2005; Swerdlow et al. 2015) showing the hereditary nature of many biomarkers. Although the contribution of unobservable factors is captured in our decomposition method, we are clearly not able to measure the contribution of these factors to the overall inequality. More generally, the set of circumstances available in our empirical application are much closer to the ones employed in the normative literature on health equity (eg., Rosa Dias and Jones, 2007). Yet, the partial observability of the circumstances is a common feature of all inequality of opportunity analyses (Ferreira and Peragine, 2015) which is hard to solve with the datasets usually available to scholars. This may lead to an underestimation of the share of *illegitimate* inequality and may bias the relationship between efforts and health by influencing the sorting of individuals in bad habits. On the other hand, our empirical application uses biomarker data. Although this complicates the use of additional variables (since biomarker measurements are available only for a random sub-sample of the full sample), it is a rare feature of health equity studies and is important to solve reporting bias issues. Moreover, our empirical application includes deprivation of the area of residence among circumstances and this partly helps to take into account also the social background of individuals. The availability of new data-sets combining a larger set of circumstances – including genetic data - and biomarkers data may contribute to relax this trade-off and add significantly to our understanding of how social and economic circumstances impact on health outcomes (Benzeval, et al. 2016).

References

Aaberge, R., Mogstad, M., Peragine, V., (2011), Measuring long-term inequality of opportunity, *Journal of Public Economics*, 95: 193-204.

Abatemarco, A., (2015), A Gini approach to inequality of opportunity: evidence from the PSID, *Empirical Economics*, 49: 1497-1519.

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 Pharmacology and Therapeutics*, 69: 89–95.

Aune, D., Giovannucci, E., Boffetta, P., Fadnes, L.T., Keum, N., Norat, T., Greenwood, D.C., Riboli, E., Vatten, L.J., Tonstad, S., (2017), Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies, *International Journal of Epidemiology*, 46: 1029–1056.

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

Benzeval, M., Kumari, M., Jones, A., (2016), How do biomarkers and genetics contribute to understanding society?, *Health Economics*, 25: 1219-1222.

Bilger, M., Carrieri, V., (2013), Health in the cities: when the neighbourhood natters more than income, *Journal of Health Economics*, 32:1-11.

Bourguignon, F., Ferreira, F.H.G. and Menendez, M., (2007) Inequality of opportunity in Brazil, *Review of Income and Wealth* 53: 585-618.

Burchardt, T., Le Grand J., (2002) Constraint and opportunity: identifying voluntary non-employment, Centre for Analysis of Social Exclusion, Paper 55, London School of Economics.

Carrieri, V., Jones, A.M., (2017), The income-health relationship 'beyond the mean': new evidence from biomarkers, *Health Economics,* 26: 937-956.

Checchi, D., Peragine, V., (2005), Regional disparities and inequality of opportunity: the case of Italy, *IZA Discussion Papers* 1874.

Checchi, D., Peragine, V., (2010), Inequality of opportunity in Italy, *Journal of Economic Inequality,* 8: 429-450.

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.

Erreygers, G., (2009), Correcting the concentration index, *Journal of Health Economics,* 28: 504-515.

Erreygers, G., Kessels, R., (2013), Regression-based decompositions of rank-dependent indicators of socioeconomic inequality in health, in Rosa Dias, P., O'Donnell,O., (eds) *Health Inequality (Research on Economic Inequality Volume 21 Chapter 9).* London: Emerald Group Publishing.

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., (2008), *Fairness, Responsibility and Welfare*, Oxford University Press: Oxford.

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, in Barros, P., McGuire T., Pauly, M. (eds.), *Handbook of Health Economics, Volume 2*: 1003-1092.

Garcia-Gomez, P., Schokkaert, E., Van Ourti, T., Bago d'Uva, T., (2015), Inequity in the face of death, *Health Economics,* 24:1348-67.

Goodman, E., McEwen, B. S., Huang, B., et al., (2005), Social inequalities in biomarkers of cardiovascular risk in adolescence, *Psychosomatic Medicine*, 67: 9-15.

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.

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

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.

Kim, B., (2016), Inequality of opportunity for healthy aging in Europe, KU Leuven Department of Economics, Discussion paper 16/20, University of Leuven.

Lefranc, A., Pistolesi, N., Trannoy, A., (2009), Equality of opportunity and luck: definitions and testable conditions, with an application to income in France, *Journal of Public Economics*, 93: 1189-1207.

Li Donni, P., Peragine, V., Pignataro, G., (2014), Ex-ante and Ex-post measurement of equality of opportunity in health: a normative decomposition, *Health Economics,* 23: 182-198.

Li Donni, P., Rodriguez, J.G., Rosa Dias, P., (2015), Empirical definition of social types in the analysis of inequality of opportunity: a latent class approach, *Social Choice and Welfare*, 44: 673-701.

Rao, V.M., (1969), Two decompositions of concentration ratio, *Journal of the Royal Statistical Society. Series A (General),* 132: 418-425.

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.

Roemer, J.E. and Trannoy, A. (2016) Equality of opportunity: theory and measurement, *Journal of Economic Literature,* 54: 1288-1332.

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.

Shorrocks, A., (1982), Inequality decomposition by factor components, *Econometrica,* 50: 193-212.

Swerdlow, D. I., Preiss, D., Kuchenbaecker, et al. (2015), HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials, *The Lancet*, 385: 351-361

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

Turk, U., Östh, J., (2017), Inequality of opportunity in Sweden: a spatial perspective, *University of Verona Department of Economics Working Paper WP 9*.

Van Doorslaer, E., Jones, A.M. (2003), Inequalities in self-reported health: validation of a new approach to measurement, *Journal of Health Economics*, 22: 61-87.

Van Ourti, T., van Doorslaer, E., Koolman, K., (2009), The effect of income growth and inequality on health inequality: theory and empirical evidence from the European panel, *Journal of Health Economics* 28: 525-539.

Wagstaff, A., van Doorslaer, E., Watanabe, N., (2003), On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam, *Journal of Econometrics,* 112: 207-223.

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 3. Decomposition Results – Cholesterol

Variables	Effort	% [a]	Direct circumstances	% [a]	Indirect circumstances	% [a]	Total	%
GINI DECOMPOSITION								
Cotinine	0.000441	**1.90**			0.000193	**0.83**		
Drinking	0.000451	**1.95**			0.000311	**1.34**		
Bad diet	0.000441	**1.90**			0.000392	**1.69**		
Total contribution of variables	0.001333	**5.76**	0.020908	**90.36**	0.000896	**3.49**	**0.0231**	**20.72**
Residuals							**0.0880**	**79.28**
Gini							**0.1116**	
VARIANCE DECOMPOSITION								
Cotinine	0.005063	**2.12**			0.002118	**0.88**		
Drinking	0.004658	**1.95**			0.002945	**1.23**		
Bad diet	0.004693	**1.96**			0.004368	**1.82**		
Total contribution of variables	0.014414	**6.03**	0.215305	**90.03**	0.009431	**3.94**	**0.2391**	**19.74**
Residuals							**0.9721**	**80.25**
Variance							**1.2112**	

[a] In percentage of explained inequality

Table 4. Decomposition Results – Glycated Haemoglobin

Variables	Effort	% [a]	Direct circumstances	% [a]	Indirect circumstances	% [a]	Total	%
GINI DECOMPOSITION								
Cotinine	0.000363	**3.31**			-0.000462	**-4.17**		
Drinking	0.000548	**4.95**			-0.000941	**-8.51**		
Bad diet	0.000142	**1.28**			0.000938	**8.47**		
Total contribution of variables	0.001057	**9.56**	0.010469	**94.65**	-0.000465	**-4.21**	**0.0110**	**21.73**
Residuals							**0.0398**	**78.27**
Gini							**0.0509**	
VARIANCE DECOMPOSITION								
Cotinine	0.002762	**5.02**			-0.002229	**-4.05**		
Drinking	0.003654	**6.64**			-0.004502	**-8.18**		
Bad diet	0.000918	**1.67**			0.004254	**7.73**		
Total contribution of variables	0.007334	**13.33**	0.050177	**91.17**	-0.0026935	**-4.50**	**0.0550**	**13.50**
Residuals							**0.3524**	**86.49**
Variance							**0.4075**	

[a] In percentage of explained inequality

Table 5. Decomposition Results – Fibrinogen

Variables	Effort	% [a]	Direct circumstances	% [a]	Indirect circumstances	% [a]	Total	%
GINI DECOMPOSITION								
Cotinine	0.004675	**19.24**			-0.000795	**-3.27**		
Drinking	0.003029	**12.47**			-0.001883	**-7.75**		
Bad diet	0.001951	**8.03**			0.003727	**15.35**		
Total contribution of variables	0.009655	**39.75**	0.013582	**55.92**	0.001049	**4.31**	**0.0242**	**19.84**
Residuals							**0.0981**	**80.16**
Gini							**0.1224**	
VARIANCE DECOMPOSITION								
Cotinine	0.016266	**20.09**			-0.002392	**-2.95**		
Drinking	0.011262	**13.91**			-0.005299	**-6.54**		
Bad diet	0.006518	**8.05**			0.006518	**12.14**		
Total contribution of variables	0.034046	**42.05**	0.044779	**55.31**	0.002140	**2.64**	**0.0809**	**18.96**
Residuals							**0.3461**	**81.04**
Variance							**0.4271**	

[a] In percentage of explained inequality

Table 6. Decomposition Results – Ill Health Index

Variables	Effort	% [a]	Direct circumstances	% [a]	Indirect circumstances	% [a]	Total	%
GINI DECOMPOSITION								
Cotinine	0.004032	**8.06**			0.0000780	**0.15**		
Drinking	0.003205	**6.41**			-0.002964	**-5.92**		
Bad diet	0.001940	**3.88**			0.005925	**11.84**		
Total contribution of variables	0.009177	**18.34**	0.037822	**75.58**	0.0025475	**6.07**	**0.0500**	**28.94**
Residuals							**0.1228**	**71.06**
Gini							**0.1729**	
VARIANCE DECOMPOSITION								
Cotinine	0.031801	**8.19**			0.000110	**0.03**		
Drinking	0.026599	**6.85**			-0.023501	**-5.99**		
Bad diet	0.017322	**4.46**			0.0415266	**10.70**		
Total contribution of variables	0.07572	**19.51**	0.293951	**75.75**	0.018387	**4.73**	**0.3880**	**27.39**
Residuals							**1.0286**	**72.61**
Variance							**1.4167**	

[a] In percentage of explained inequality

ONLINE APPENDIX

I. OTHER DECOMPOSITIONS

The Inequality of Opportunity Gini

The decomposition of G^{IOp} follows the same logic described in equations (5)-(10). The benchmark uses the average health of the worst-off type ($\overline{H^w}$). 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 (8) and (9), the decomposition of G^{IOp} can be expressed as follows :

$$\begin{aligned} G^{IOp} = \left(\frac{2}{N\overline{H^w}}\right) \sum_\tau \sum_{i\in\tau} (\alpha_\tau - \overline{\alpha^w})(R_i - \tfrac{1}{2}) + \\ \left(\tfrac{2}{N\overline{H^w}}\right) \textstyle\sum_\tau \sum_{i\in\tau} (\bar{B}_\tau - \overline{B^w})(R_i - \tfrac{1}{2}) + \\ \left(\tfrac{2}{N\overline{H^w}}\right) \textstyle\sum_\tau \sum_{i\in\tau} \ (B_i - \overline{B_\tau})(R_i - \tfrac{1}{2}) + \\ \left(\frac{2}{N\overline{H^w}}\right) \sum_\tau \sum_{i\in\tau} u_i^\tau (R_i - \tfrac{1}{2}) \end{aligned} \tag{11}$$

The terms in equation (11) follow the same logic of those in equation (10) 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 effort*, while the final term measures the *contribution of residual factors*.

Absolute Gini

The absolute Gini is given by:

$$AG = 2Cov(H_i, R_i) \tag{12}$$

Regression-based decomposition of this index does not require an assumption of weighting function ignorability (as the weighting function is a constant). The decomposition becomes:

$$\begin{aligned} AG = \left(\frac{2}{N}\right) \sum_\tau \sum_{i\in\tau} (\alpha_\tau - \bar{\alpha})(R_i - \tfrac{1}{2}) + \\ \left(\tfrac{2}{N}\right) \textstyle\sum_\tau \sum_{i\in\tau} (\bar{B}_\tau - \bar{B})(R_i - \tfrac{1}{2}) + \\ \left(\tfrac{2}{N}\right) \textstyle\sum_\tau \sum_{i\in\tau} \ (B_i - \overline{B_\tau})(R_i - \tfrac{1}{2}) + \\ \left(\tfrac{2}{N}\right) \textstyle\sum_\tau \sum_{i\in\tau} \ u_i^\tau (R_i - \tfrac{1}{2}) + \end{aligned} \tag{13}$$

The percentage contributions from this decomposition of the absolute Gini are identical to those for the relative Gini. So, as long as the percentagewise decomposition is the main focus of attention – as it is here - it is clear that weighting function ignorability is not a concern.

Variance

In simplified form, Shorrock's (1982) decomposition relies on the fact that if

$$y = x + z$$

then

$$var(y) = cov(x, y) + cov(z, y)$$

Given (3), (6) and (7) we have:

$$H_i - \overline{H} = \alpha_\tau - \bar{\alpha} + B_i - \bar{B} + u_i^\tau \tag{14}$$

which can be expanded to give:

$$H_i - \overline{H} = (\alpha_\tau - \bar{\alpha}) + (\bar{B}_\tau - \bar{B}) + (B_i - \overline{B_\tau}) + u_i^\tau \tag{15}$$

Using the Shorrocks decomposition this gives:

$$Var(H) = cov(\alpha_\tau - \bar{\alpha}, H) + cov\big((\bar{B}_\tau - \bar{B}), H\big) + cov\big((B_i - \bar{B}_\tau), H\big) + cov(u_i^\tau, H) \tag{16}$$

II. GLOSSARY OF BIOMARKERS

Cotinine
Cotinine is the predominant metabolite of nicotine and it is a quantitative indicator of active smoking. Cotinine levels greater than or equal to 12ng/ml usually identify active smoking with high sensitivity (96.7%; Jarvis et al. 2008).

Total Cholesterol
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 cut-points 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.

III. SUPPLEMENTARY RESULTS

Table A.1 Distribution of types

Types	Types Definition [a]	Biomarkers (Sample Size)			
		Cholesterol	Glycated Haemoglobin	Fibrinogen	Ill Health Index
1	LE,F,<59, HD	449	578	123	122
2	LE,F,<59, LD	1016	1268	249	247
3	LE, F, 60-79, HD	389	391	113	109
4	LE, F, 60-79, LD	556	551	146	140
5	LE, F, 80+, HD	140	139	34	34
6	LE, F, 80+, LD	109	109	20	20
7	LE, M, <59, HD	438	639	152	147
8	LE, M, <59, LD	752	1,090	190	186
9	LE, M, 60-79, HD	386	403	122	117
10	LE, M, 60-79, LD	439	459	125	123
11	LE, M, 80+, HD	101	102	25	24
12	LE, M, 80+, LD	127	126	23	23
13	ME, F, <59, HD	94	110	32	29
14	ME, F, <59, LD	422	485	96	96
15	ME, F, 60-79,HD	231	229	61	57
16	ME, F, 60-79, LD	498	497	115	113
17	ME, F, 80+,HD	94	93	26	21
18	ME, F, 80+,LD	157	153	28	28
19	ME, M, <59, HD	127	171	32	31
20	ME, M, <59, LD	521	659	115	115
21	ME, M, 60-79, HD	222	231	64	64
22	ME, M, 60-79, LD	491	505	124	118
23	ME, M, 80+, HD	132	131	31	31
24	ME, M, 80+, LD	211	206	47	44
25	HE, F, <59, HD	78	83	22	21
26	HE, F, <59, LD	375	411	82	81
27	HE, F, 60-79, HD	200	202	65	62
28	HE, F, 60-79, LD	617	611	133	129
29	HE, F, 80+, HD	75	75	7	7
30	HE, F, 80+, LD	134	134	26	26
31	HE, M, <59, HD	91	119	17	17
32	HE, M, <59, LD	485	605	111	109
33	HE, M, 60-79, HD	171	177	62	62
34	HE, M, 60-79, LD	559	575	144	139
35	HE, M, 80+, HD	79	72	12	11
36	HE, M, 80+, LD	130	127	21	21
	Observations	11096	12516	2795	2724

[a] LE= Qualification below nvq3GCE A level, ME=Nvq 3/GCE A level, HE= nvq4/nvq5/degree or equivalent; F=Female, M=Male; <59=Born before 1959, 60-79=Born between 1960 and 1979, 80+= Born after 1980; LD= Living in a low deprived area, HD= Living in a high deprived area.

Table A.2 Means of effort variables by type

	Type	Cotinine	Drinking	Bad Diet
1	LE,F,<59, HD	72.03	11.01	5.74
2	LE,F,<59, LD	33.94	10.45	5.17
3	LE, F, 60-79, HD	129.17	17.84	6.53
4	LE, F, 60-79, LD	73.19	14.76	6.16
5	LE, F, 80+, HD	99.53	16.37	6.58
6	LE, F, 80+, LD	83.26	14.20	6.22
7	LE, M, <59, HD	107.20	23.45	5.99
8	LE, M, <59, LD	53.73	21.33	5.28
9	LE, M, 60-79, HD	159.79	34.44	6.74
10	LE, M, 60-79, LD	89.68	33.01	6.41
11	LE, M, 80+, HD	110.89	28.68	6.85
12	LE, M, 80+, LD	85.00	29.35	6.63
13	ME, F, <59, HD	61.77	12.43	4.88
14	ME, F, <59, LD	26.83	12.31	4.47
15	ME, F, 60-79,HD	67.57	16.42	5.98
16	ME, F, 60-79, LD	50.13	15.59	5.32
17	ME, F, 80+,HD	72.23	20.31	5.71
18	ME, F, 80+,LD	30.95	18.78	5.85
19	ME, M, <59, HD	81.05	26.09	5.37
20	ME, M, <59, LD	37.85	22.91	5.11
21	ME, M, 60-79, HD	99.48	32.77	6.16
22	ME, M, 60-79, LD	61.12	29.66	5.93
23	ME, M, 80+, HD	61.79	38.52	6.28
24	ME, M, 80+, LD	46.94	36.20	5.94
25	HE, F, <59, HD	50.66	14.49	4.38
26	HE, F, <59, LD	15.68	14.80	4.29
27	HE, F, 60-79, HD	34.20	14.32	5.04
28	HE, F, 60-79, LD	14.66	14.42	4.81
29	HE, F, 80+, HD	23.11	14.03	4.60
30	HE, F, 80+, LD	17.04	17.05	5.06
31	HE, M, <59, HD	57.09	24.94	4.69
32	HE, M, <59, LD	19.10	22.77	4.55
33	HE, M, 60-79, HD	51.54	27.54	5.33
34	HE, M, 60-79, LD	29.80	24.44	5.29
35	HE, M, 80+, HD	54.42	35.90	5.20
36	HE, M, 80+, LD	30.42	30.95	5.04

Table A.3. Means of biomarkers by type

	Type	Cholesterol	Glycated Haemoglobin	Fibrinogen	Ill Health Index
1	LE,F,<59, HD	6.13	5.77	3.16	4.47
2	LE,F,<59, LD	6.16	5.74	3.17	4.43
3	LE, F, 60-79, HD	5.28	5.43	3.12	3.63
4	LE, F, 60-79, LD	5.20	5.36	2.81	3.36
5	LE, F, 80+, HD	4.60	5.24	2.87	3.01
6	LE, F, 80+, LD	4.51	5.23	2.67	2.66
7	LE, M, <59, HD	5.65	5.90	3.21	4.38
8	LE, M, <59, LD	5.71	5.85	3.17	4.37
9	LE, M, 60-79, HD	5.61	5.51	2.83	3.63
10	LE, M, 60-79, LD	5.66	5.52	2.81	3.78
11	LE, M, 80+, HD	4.70	5.33	2.39	2.52
12	LE, M, 80+, LD	4.50	5.31	2.55	2.88
13	ME, F, <59, HD	6.16	5.78	3.33	4.54
14	ME, F, <59, LD	6.20	5.70	3.16	4.43
15	ME, F, 60-79,HD	5.12	5.37	3.00	3.40
16	ME, F, 60-79, LD	5.20	5.33	2.86	3.32
17	ME, F, 80+,HD	4.71	5.22	2.98	3.26
18	ME, F, 80+,LD	4.70	5.26	2.80	3.21
19	ME, M, <59, HD	5.72	5.81	3.08	4.08
20	ME, M, <59, LD	5.78	5.78	3.05	4.37
21	ME, M, 60-79, HD	5.62	5.45	2.68	3.33
22	ME, M, 60-79, LD	5.66	5.45	2.72	3.59
23	ME, M, 80+, HD	4.75	5.26	2.48	2.77
24	ME, M, 80+, LD	4.63	5.25	2.58	2.84
25	HE, F, <59, HD	6.11	5.69	3.11	4.17
26	HE, F, <59, LD	6.12	5.65	3.14	4.44
27	HE, F, 60-79, HD	5.13	5.36	2.83	3.27
28	HE, F, 60-79, LD	5.13	5.31	2.72	3.18
29	HE, F, 80+, HD	4.76	5.16	3.29	3.32
30	HE, F, 80+, LD	4.72	5.22	2.93	3.22
31	HE, M, <59, HD	5.62	5.82	3.06	4.65
32	HE, M, <59, LD	5.77	5.69	2.92	4.25
33	HE, M, 60-79, HD	5.39	5.36	2.75	3.32
34	HE, M, 60-79, LD	5.50	5.39	2.65	3.45
35	HE, M, 80+, HD	4.69	5.34	2.46	2.48
36	HE, M, 80+, LD	4.76	5.23	2.60	3.00
	Observations	11096	12516	2795	2724

Table A.4 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	-0.001***	0.000	-0.001	0.001	-0.001	0.001***	0.000	0.000	0.000	-0.001
Drinking	0.001	0.003	0.001	-0.002	-0.001	0.002	0.002	0.004***	0.002	0.001	0.002	-0.001	-0.006	0.002	0.002	0.005***	-0.004	0.000
Bad diet	0.017	-0.003	0.003	-0.014	-0.027	-0.052	0.012	0.058***	0.032	-0.015	0.050	-0.015	0.016	0.025	0.026	0.052***	0.008	0.005
Constant	5.991***	6.134***	5.200***	5.316***	4.875***	4.883***	5.504***	5.255***	5.431***	5.724***	4.408***	4.579***	6.199***	6.000***	4.873***	4.791***	4.690***	4.711***
Obs.	449	1016	389	556	140	109	438	752	386	439	101	127	94	422	231	498	94	157

Cont'd

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Cotinine	0.001*	-0.000	-0.000	0.000	-0.002**	0.000	-0.000	0.001	-0.000	0.001***	0.002	-0.001	0.001*	-0.000	0.000	0.001	0.000	0.000
Drinking	0.001	0.001	0.000	0.004**	0.001	0.001	0.004	-0.003	0.004	0.002	-0.003	0.002	0.004	0.003*	0.001	0.001	0.001	-0.001
Bad diet	-0.046	0.028	0.018	0.034	-0.045	-0.003	-0.066	-0.006	-0.021	0.021	0.016	0.055	0.009	-0.014	0.032	0.032*	0.017	0.033
Constant	5.898***	5.611***	5.518***	5.310***	5.154***	4.613***	6.395***	6.187***	5.180***	4.965***	4.685***	4.366***	5.378***	5.777***	5.189***	5.266***	4.537***	4.579***
Obs	127	521	222	491	132	211	78	375	200	617	75	134	91	485	171	559	79	130

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

Table A.5 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	0.001***	0.000	0.000	-0.000	-0.000	0.000	0.000	0.001***	-0.000	0.000	0.000*	0.000	0.000	0.000	0.000
Drinking	-0.005***	-0.004***	0.002	-0.002***	-0.001	-0.000	-0.001	-0.004***	-0.001	0.001	-0.001	-0.001	-0.007*	-0.004***	-0.001	-0.001	-0.001	0.000
Bad diet	0.004	0.006	0.001	-0.012	0.001	-0.011	-0.005	0.020*	0.009	0.022*	0.013	0.014	0.042	0.019	0.003	0.018**	-0.003	-0.011
Constant	5.820***	5.747***	5.373***	5.439***	5.222***	5.285***	5.983***	5.825***	5.445***	5.314***	5.175***	5.268***	5.640***	5.654***	5.373***	5.230***	5.243***	5.320***
Obs.	578	1268	391	551	139	109	639	1090	403	459	102	126	110	485	229	497	93	153

Cont'd

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Cotinine	-0.001	0.000	0.000	0.000	0.000	0.001***	0.000	0.000	0.000	0.000*	0.000	0.001	0.001*	0.001***	0.000	0.001***	0.004***	0.000
Drinking	-0.001	-0.003**	-0.001	-0.001	0.000	0.000	-0.004	-0.002	-0.002	-0.002**	-0.000	-0.003**	-0.006*	-0.002**	-0.000	-0.001	0.003	-0.001
Bad diet	0.009	0.002	0.037**	-0.010	0.003	0.010	0.001	-0.013	0.001	-0.015**	-0.012	0.006	-0.006	-0.016	0.003	-0.005	-0.047	-0.006
Constant	5.827***	5.820***	5.197***	5.532***	5.222***	5.140***	5.759***	5.739***	5.381***	5.419***	5.223***	5.223***	5.959***	5.805***	5.345***	5.414***	5.411***	5.287***
Obs	171	659	231	505	131	206	83	411	202	611	75	134	119	605	177	575	72	127

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

Table A.6 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	0.001	0.000	-0.000	0.001	-0.001	0.001	0.001**	0.000*	0.001**	0.001***	0.002**	-0.000	-0.000	0.000	-0.000	0.000	0.001	-0.001
Drinking	-0.005	-0.009***	0.003	-0.011***	-0.016**	0.005	-0.001	-0.003	-0.000	0.001	-0.004	0.000	-0.014*	-0.011***	0.016**	-0.004	0.005	-0.001
Bad diet	0.030	0.006	0.043	-0.006	0.037	-0.154*	-0.016	0.021	0.079***	-0.024	0.037	0.009	0.035	0.021	-0.056*	-0.023	-0.156**	0.025
Constant	2.982***	3.228***	2.809***	2.978***	2.917***	3.746***	3.249***	3.070***	2.162***	2.826***	1.968***	2.516***	3.327***	3.174***	3.241***	3.039***	3.944***	2.659***
Obs.	123	249	113	146	34	20	152	190	122	125	25	23	32	96	61	115	26	28

Cont'd

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Cotinine	0.000	0.001***	0.000	0.001***	0.001	0.000	0.001***	0.001	0.003***	0.001**	-0.011	-0.005	-0.001	0.001**	0.002**	0.001**	0.002	-0.000
Drinking	-0.005	0.002	0.002	-0.000	0.000	0.004*	-0.003	-0.003	-0.004	-0.001	-0.020	0.003	-0.010*	-0.002	0.010***	-0.001	-0.004	-0.004
Bad diet	0.059	-0.007	0.002	0.005	-0.080**	-0.016	0.076*	-0.069**	-0.019	0.013	0.046	0.009	0.054	0.001	-0.018	0.016	-0.039	-0.153**
Constant	2.745***	2.955***	2.600***	2.614***	3.028***	2.561***	2.670***	3.518***	2.895***	2.631***	3.518**	2.820***	3.182***	2.926***	2.522***	2.547***	2.822**	3.777***
Obs	32	115	64	124	31	47	22	82	65	133	7	26	17	111	62	144	12	21

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

Table A.7 Regressions by type – Ill Health Index

	Type																	
Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Cotinine	0.001	0.001	-0.001	0.001	-0.002	0.001	0.000	0.000	0.000	0.001**	0.000	0.000	-0.000	0.001	0.001	-0.001	-0.001	-0.002
Drinking	-0.016**	-0.006	0.007*	-0.019***	-0.025**	0.020	-0.001	-0.005	0.001	0.002	0.007	-0.002	-0.032***	-0.019**	0.014	-0.001	-0.009	0.002
Bad diet	0.114**	0.024	0.028	-0.014	0.044	-0.189	-0.051	0.083**	0.092*	-0.060	0.046	0.046	-0.020	0.059	-0.082*	0.033	-0.025	-0.014
Constant	3.829***	4.332***	3.394***	3.688***	3.222***	3.898***	4.753***	3.963***	2.877***	3.998***	2.009***	2.546***	4.955***	4.283***	3.721***	3.173***	3.602***	3.304***
Obs.	122	247	109	140	34	20	147	186	117	123	24	23	29	96	57	113	21	28

Cont'd

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Cotinine	0.000	0.002**	0.000	0.001**	0.003	0.001	0.002**	0.001	0.005**	0.002***	-0.016*	-0.014	-0.002	0.003***	0.002**	0.002***	0.001	-0.002
Drinking	-0.007	-0.002	-0.002	-0.000	-0.000	0.002	-0.002	-0.007	-0.008	0.001	-0.014	0.004	-0.021	-0.004	0.010**	-0.005*	-0.009	-0.000
Bad diet	-0.045	0.038	-0.017	0.029	-0.035	-0.030	0.100	-0.131**	-0.046	0.039	-0.097	-0.066	-0.102	-0.011	0.015	0.025	0.026	-0.114
Constant	4.449***	4.086***	3.480***	3.318***	2.930***	2.933***	3.422***	5.197***	3.500***	2.865***	4.484***	3.617***	5.892***	4.329***	2.882***	3.318***	2.605	3.815***
Obs	31	115	64	118	31	44	21	81	62	129	7	26	17	109	62	139	11	21

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

IV. ROBUSTNESS CHECKS

In this section we report a number of robustness checks and sensitivity analyses of our decomposition results under alternative definition of types and/or a different definition of circumstances and effort variables. All the results discussed in this Section are not reported but they are available upon request.

One potential concern with the results shown in Section 4 might be the relatively small sub-samples by type for some biomarkers, especially fibrinogen and the ill-health index. This might artificially increase the heterogeneity of coefficients across types and be capturing sampling variation across these relatively small sub-samples. In order to check for this possibility, we experimented with the definitions of types by considering different combinations of circumstances variables. We first deleted residential status among the circumstances variables and this leads to 18 types (instead of the 36 types actually employed) and a significant increase in sub-sample size. Secondly, we split education in two categories (below/above compulsory schooling) using cohort and gender as other circumstances variables. This leads to 12 types and a further gain in the average sub-sample size. Interestingly, these experiments affect some of the regression results but leave the sign and the magnitude of the decomposition results substantially unchanged.

As a second check, we considered, separately, other additional circumstance variables available in our dataset for the definition of types. First, we experimented with parental smoking status, namely whether or not parents smoked when the respondent was a child and we deleted residential status from the set of circumstances variables. Also in this case, our decomposition results are substantially unchanged. Second, we considered the possibility of including the ethnicity of the respondents as additional circumstance variable. However, we found that the share of non-white individuals in the older cohorts and in the higher levels of education was too small, e.g. less than 5 individuals for some types in the case of fibrinogen. This would lead to a very imprecise decomposition of inequality and may not bring substantial differences to our main results, because of a strong correlation of ethnicity with the other circumstances (education and cohort). Similarly, other information collected in the HSE is ruled out due to low sample sizes: information on whether an individual was born prematurely and information on whether parents died and their cause of death are available only for a few waves and produce a very limited sample sizes.

Finally, we also experimented with the definition of effort variables. First, we include whether the individuals take medications prescribed by the doctor among the effort factors. This variable is useful to control for the fact that medications might be actually prescribed by the doctor in response to adverse biomarker scores thus potentially affecting our decomposition results. When included among the effort variables, we find a significant and positive association between medications and biomarkers and a significant contribution of this variable among the decomposition terms. However, the inclusion of this additional effort variable does not affect the pattern of decomposition. We also consider the inclusion of BMI, the type of work (i.e. job strenuousness) and sport activity variables among the effort variables as they may affect significantly biomarkers scores. However, the inclusion of these variables in an IOp framework is problematic for a number of reasons. For what concerns BMI, it is non-monotonic in efforts since both being under-weight and over-weight are health problems. Secondly, IOp requires partitioning factors into circumstances and efforts, while BMI may reflect partly genetics, which is typically regarded as a circumstance, as well as 'lifestyle choices' that may be seen as an effort. A similar argument applies to the

type of work as it is partly under the individual control (thus an effort factor) but also strongly dependent on education and skills which are partly related to the socio-economic background of the individual (thus a circumstance). An additional concern with the use of BMI comes from the scaling of the effort variables in our decomposition method. Specifically, a BMI equals to 0 does not have practical relevance. The inclusion of physical activity among the effort variables is less contentious. Unfortunately, information on sporting activity is available only in selected waves of the HSE and the measures are not entirely comparable across waves. For instance, using the variable which is repeated most often (the average hours of sport in a week) the number of observations available would drop by around 70% for the analysis of cholesterol, by 60% for glycated haemoglobin and by 100% for the remaining biomarkers. This would be problematic for our empirical approach which is based on estimating separate regressions of health outcomes on effort for each of the sub-samples defined by types