Article:

Carrieri, Vincenzo, Davillas, Apostolos and Jones, Andrew Michael orcid.org/0000-0003-4114-1785 (2020) A latent class approach to inequity in health using biomarker data. Health Economics. p. 808. ISSN 1057-9230

https://doi.org/10.1002/hec.4022

A latent class approach to inequity in health using biomarker data

Vincenzo Carrieri¶
Apostolos Davillas†
Andrew M Jones‡

Abstract

We adopt an empirical approach to analyse, measure and decompose Inequality of Opportunity (IOp) in health, based on a latent class model. This addresses some of the limitations that affect earlier work in this literature concerning the definition of types, such as partial observability, the *ad hoc* selection of circumstances, the curse of dimensionality and unobserved type-specific heterogeneity that may lead to biased estimates of IOp. We apply our latent class approach to measure IOp in allostatic load, a composite measure of biomarker data. Using data from *Understanding Society* (UKHLS), we find that a latent class model with three latent types best fits the data, with the corresponding types characterised in terms of differences in their observed circumstances. Decomposition analysis shows that about two-thirds of the total inequality in allostatic load can be attributed to the direct and indirect contribution of circumstances and that the direct contribution of effort is small. Further analysis conditional on age-sex groups reveals that the relative (percentage) contribution of circumstances to the total inequalities remains mostly unaffected and the direct contribution of effort remains small.

Keywords: equality of opportunity; health equity; biomarkers; finite mixture models; latent class models; decomposition analysis

JEL codes: C1, D63, I12, I14.

Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service. Andrew Jones acknowledges funding from the Leverhulme Trust Major Research Fellowship (MRF-2016-004). The funders, data creators and UK Data Service have no responsibility for the contents of this paper.
¶ Department of Law, Economics and Sociology, "Magna Graecia" University of Catanzaro; IZA (Institute of Labor Economics); RWI-Research Network.
† Health Economics Group, University of East Anglia and Institute for Social and Economic Research, University of Essex.
‡ Department of Economics and Related Studies, University of York; Centre for Health Economics, Monash University

1 Introduction

Based on Roemer's (1998, 2002) influential formalisation of the concept, a large body of empirical research has dealt with the assessment of inequality of opportunity (IOp) for a variety of measures of well-being. The IOp literature argues that the egalitarian framework does not necessarily dictate equality of the distribution of outcomes *per se* but emphasises the role of individual responsibility in defining a "fair" distribution. Early contributions to the IOp literature have focused mainly on income (see Ramos and van de Gaer (2016) and Roemer and Trannoy (2016) for reviews). More recently, a growing literature has addressed the measurement of IOp in other relevant dimensions of individual well-being such as education (Ferreira and Gignoux, 2013) and health (e.g., Rosa Dias, 2009; Rosa Dias, 2010; Trannoy et al., 2010; Jusot, et al., 2013; García-Gómez et al., 2015; Deutsch et al., 2018; Davillas and Jones, 2020).

This literature separates the factors associated with an outcome of interest into two components: 'circumstances', which are not under individual responsibility and are viewed as an illegitimate or unfair source of inequality, and 'efforts' for which, to some extent, individuals are held responsible and that are viewed as a legitimate source of inequality. Following Roemer (1998, 2002), the IOp literature often defines types as a group of individuals who share the same set of circumstances, such as parental background and early life circumstances (e.g., Aaberge et al., 2011; Carrieri and Jones, 2018; Fleurbaey and Peragine, 2013; Ramos and van de Gaer, 2016; Trannoy et al., 2010). In the context of health equity, Fleurbaey and Schokkaert (2009, 2012) take a broader perspective that uses the responsibility cut to distinguish factors that are seen as fair sources of inequality of outcomes and those that are seen as unfair, with the health variations attributed to the latter is regarded as health inequity. In this study, we adopt a social perspective and draw on the socio-legal context of the UK health system to define the sources of the unfair variation.

A key empirical challenge in these analyses is the definition of types. It is difficult to devise a criterion to make the Roemer model operational, especially because the original model does not provide practical guidance for either the number or the combination of circumstances that should be used to define social types (Li Donni et al., 2015). This implies that a large part of the existing empirical research in IOp may have a number of limitations. First, researchers may observe only a limited set of circumstances, with the partial observability of the circumstances is often a common feature of IOp studies (see Brunori et al., 2019 and Li Donni et al., 2015 for relevant discussions). This may lead to an underestimation of the share of illegitimate inequality. Second, researchers often rely on *ad hoc* definitions of types according to exposure to a small number of circumstances which, although they may be guided by the norms and conventions of the society being analysed, are more or less arbitrarily selected by the researcher (Li Donni et al., 2015, Brunori et al., 2019). Third, the combination of selected circumstances into types may result in a trade-off between the number of types and the sample size for each type. For example, the high correlation between different measures of parental socioeconomic status can make it hard to define clear cut and mutually exclusive categories, resulting in types with few observations, which may lead to overestimates in the measurement of IOp due to sampling variance in the distribution of type means (Brunori et al., 2019). Researchers often address these problems by using a limited number of circumstances or an arbitrary aggregation of socioeconomic categories. The curse of dimensionality may imply severe limitations given that stochastic dominance tests, often employed as a first stage to identify the presence of IOp, are highly sensitive to the choice of circumstance variables, as are results from analyses that involve separate regressions by type (e.g., Bourguignon et al., 2007; Carrieri and Jones, 2018; Garcia-Gomez et al., 2015). Beyond nonparametric analysis, reliability of parametric IOp estimates may also require a sufficient number of observations in each category to characterize circumstances (Brunori et al., 2019).

Building on the work of Bago d'Uva et al. (2009) on horizontal inequity in health care, Balia and Jones (2011) on IOp in mortality, and Li Donni et al. (2015) on IOp in life satisfaction, we use an semiparametric empirical approach to quantify and decompose IOp in health based on latent class models. We employ data from *Understanding Society: the UK Household Longitudinal Study (UKHLS)*, a nationally representative study that allows for objectively measured nurse-collected and blood-based biomarker data, and for a rich set of circumstance and effort variables. Specifically, we apply finite mixture models (FMMs), a semiparametric approach to model unobserved heterogeneity regarding type membership, which, unlike most of the existing IOp studies, avoids *a priori* grouping of individuals into types. Instead, FMMs are a semiparametric method to classify individuals into latent classes (types), and allows the likelihood of latent class membership to be a function of the set of observed circumstance variables. This analysis allows us to select the optimal number of latent classes (types) that are consistent with the data generation process.

A potential disadvantage of defining social types in terms of latent classes is that they are treated as a "black box", which may be hard to interpret and to assign a normative significance. We therefore augment our FMM analysis with post-estimation analysis to help characterise the latent types in terms of the combination of observed circumstances that each of them may reflect, and classify individuals into the different latent types based on the estimated posterior probabilities of class membership.

Capitalising on this useful feature of FMMs to classify individuals into latent types, we adopt and extend a recently developed decomposition technique to decompose health inequality (Carrieri and Jones, 2018). This analysis allows us to decompose total inequality in health into the direct contribution of circumstances, their indirect contribution via the heterogenous association of efforts with health by type and the direct contribution of efforts themselves. To the best of our knowledge, this is the first study that combines the inequality decomposition analysis of Carrieri and Jones (2018) with the FMM semiparametric technique to address the curse of dimensionality. This curse of dimensionality is likely to be a problem for any approach to the measurement of IOp that defines types "nonparametrically". The "nonparametric" approaches define types by using a unique combination of the values of the circumstance variables and, then, condition their analysis on these types. The curse of dimensionality arises as the number of types can become prohibitively large even with relatively few circumstance variables; for example, with just five circumstance variables, each having five categories, there would be 5^5=3,125 unique types; our latent class approach helps to reduce the dimensionality of this problem.

Specifically, by extending FMM analysis to decompose health inequality and identify the role of IOp, our study offers a number of advantages and contributions compared to earlier work in this literature concerning the definition of types. Our analysis allows the optimal number of types and the particular combination of circumstances that are used to define each type to be determined by our model and reflect the data generation process. This avoids arbitrary combinations of circumstance variables to define types or the use of an excessive number of types that may impose upward bias in the IOp measurement (Brunori et al., 2019). The FMM methodology is also helpful here since it accounts for unobserved type-specific heterogeneity in the sense of exploring differences in the association between efforts and the health outcome by latent type. Dealing with unobserved heterogeneity regarding type membership and simultaneously allowing for heterogeneous effort-health outcome associations by types is of critical importance for measuring IOp and better understanding its underlying sources.

Finally, this paper further contributes to the health equity literature by being one of the few studies that is not based on self-reported measures to proxy individual health.[1] We use a composite cardinal biological measure that captures several

[1] Self-assessed health (SAH), one of the most popular self-reported health measures, is an inherently categorical and ordinal measure and may be subject to misreporting and is associated with comparability problems at both the individual level and between countries (eg., Bago d'Uva et al. 2008). This reporting bias has been shown to vary systematically with a number of socioeconomic characteristics that are often used to explore health inequalities, which raises doubts about the robustness of studies based on self-reported health indicators (e.g., Bago d'Uva et al., 2009; Crossley and Kennedy, 2002). More fundamentally, the ordinal scaling of SAH is not compatible with the majority of the inequality indices that can be used as they require cardinal outcomes. However, recent work by Bond and Lang (2019) highlights that any attempts to cardinalize ordinal data

health dimensions, spanning adiposity, blood pressure, inflammation, blood sugar levels and cholesterol levels. Similar measures are used to capture so-called allostatic load and are considered as measures of "wear and tear" of the body that accumulates as individuals are exposed to chronic psychosocial stressors (Davillas and Pudney, 2017; Howard and Sparks, 2016; Seeman et al., 2004). As such, allostatic load is cardinal health measure that is ideal for the purpose of the measurement of IOp because it captures physiological responses that are associated with stress and the process through which economic and social circumstances may get "under the skin" across the lifespan (McEwen, 2015; Seeman et al., 2004).

We find that a latent class model with three unobserved types provides the best fit with our data. The profiles of these types can be characterised in terms of differences in their observed demographic and parental circumstances. After classifying individuals into classes using modal assignments, post-estimation decomposition analysis shows that about 50% of the total inequality in our composite health measure (allostatic load) is attributed to the direct contribution of demographic and parental circumstances. Circumstances exert an indirect contribution to the total inequalities of around 13%, though differences in the association between our effort variables and allostatic load across types. This indicates that about two thirds of the total inequality may be attributed to circumstances. However, the direct contribution of efforts is much less important, having a contribution of around 3%. Further decomposition analysis conditional on selected age-sex groups reveals that, although the observed variation in the total inequality in allostatic load differs, the relative (percentage) contribution of circumstances to the total inequalities remains similar and the direct contribution of observed efforts remains small for all groups.

2 Methods

Following the seminal work of Roemer (1998, 2002), the IOp literature assumes a responsibility cut by which factors associated with individual attainments can be grouped into two categories: a) effort factors, for which individuals should be held partially responsible, and b) circumstances which are beyond individuals' control. In the case of health, following the IOp literature (e.g., Carrieri and Jones, 2018, Jusot et al., 2013, Rosa Dias, 2010), a generalised health production function for individual health outcomes (h_i) can be defined as a function of a vector of circumstances c_i and of efforts e_i. Assuming that circumstances are not affected by

may impose significant complications given the sensitivity of empirical results drawn from ordinal data to the scaling imposed on it.

efforts, while efforts may be influenced by circumstances (Roemer, 1998, 2002), we can write:

$$h_i = h(c_i, e(c_i, v_i), u_i) \tag{1}$$

where v_i and u_i are unobserved error terms which capture the random variation in the realised outcomes. This reflects the fact that observed realisations of health outcomes are inherently random, sometimes labelled as 'luck' in the IOp literature (Lefranc et al., 2009; Lefranc and Trannoy, 2017). To be specific, v_i represents random variation in effort that is independent of c and u_i represents random variation in the outcome that is independent of c and e. The latent class specification we propose below is used to model the conditional density function $f(h_i|\boldsymbol{c_i}, \boldsymbol{e_i})$ that is implied by equation (1).

A fundamental feature of the Roemer 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 itself. This implies that, in addition to assuming a partitioning between c and e, the IOp model assumes that effort is a function of circumstances, with circumstances being pre-determined. Effort, therefore, mediates the relationship between circumstances and outcomes, and it is meaningful to consider the direct and the indirect contribution of circumstances to the inequality in outcomes. One of the strengths of our FMM specification and associated decomposition analysis, is that it allows us to explore the type-specific unobserved heterogeneity in the association between our health measure and efforts and identify the direct and indirect role of circumstances on shaping inequalities in our health outcome.

Researchers interested in quantifying IOp in outcomes (including health), typically define social types, i.e., groups of individuals who share exposure to the same circumstances, and then measure IOp between these types (e.g., Aaberge et al., 2011; Carrieri and Jones, 2018; Fleurbaey and Peragine, 2013; Ramos and Van de Gaer, 2016; Trannoy et al., 2010). Roemer (2002) defines social types consisting of individuals who share exposure to the same set of circumstances. Although the theoretical framework for the concept of types is well developed, implementation in applied work is less straightforward. As discussed in the introduction, types are often defined in an *ad hoc* way in empirical work and they are partially observable to researchers (Li Donni et al., 2015).

In this context, latent class or, specifically, FMMs offer a number of important advantages (Bago d'Uva and Jones, 2009; Bago d'Uva et al., 2009; Balia and Jones, 2011; Li Donni et al, 2015). FMMs provide an intuitive representation of unobserved heterogeneity that may exist in the data by classifying the population into a parsimonious number of latent classes. Parsimony is important as the nonparametric approach to IOp, that defines a separate dummy variable for each unique combination of values of observed circumstance variables, is likely to suffer from a curse of dimensionality. In the FMM specification the prior probabilities of membership of the latent classes can be parameterized to depend on observed circumstance variables, and the latent classes can be interpreted as unobserved types in the context of the IOp framework. Additionally, FMMs are particularly flexible because they do not require the researcher to assume, *ex-ante*, the number of latent classes, nor to provide any *a-priori* grouping based on observed circumstance variables. Another advantage of FMMs is that they are semiparametric and do not require distributional assumptions for the mixing distribution.

To put our latent class FMM specification, and its relationship to the decomposition proposed by Carrieri and Jones (2018), in context, we begin with a simple linear parametric specification of the ex post model implied by equation (1). For clarity of exposition, and without loss of generality, assume that there is only one circumstance and one effort variable. Then the model consists of an equation for the health outcome as a function of observed circumstance and effort variables and the error term:

$$h_i = \beta_0 + \beta_1 c_i + \beta_2 e_i + u_i \tag{2}$$

and an equation for effort as a function of the circumstance variable and the error term:

$$e_i = \gamma_0 + \gamma_1 c_i + v_i \tag{3}$$

In equation (2), $\beta_1 c_i$ is the direct contribution of circumstances to the outcome and $\beta_2 e_i$ is the direct contribution of effort. Estimates of (2) can therefore provide a decomposition of IOp into these components.

Equation (3) can be substituted into (2) to solve for the reduced form:

$$h_i = (\beta_0 + \beta_2\gamma_0) + (\beta_1 + \beta_2\gamma_1)c_i + (u_i + \beta_2 v_i) = \alpha_0 + \alpha_1 c_i + \varepsilon_i \tag{4}$$

The coefficient on c_i is the total contribution of circumstances, i.e. the sum of direct and indirect contributions. This total contribution can be obtained directly by estimating the reduced form (eq. 4). Alternatively, it can be obtained by a two-step procedure that replaces the observed circumstances in equation (2) with the error term from equation (3) (e.g. Bourguignon et al., 2007; Jusot et al., 2013). This two-step approach provides a link to the decomposition proposed by Carrieri and Jones (2018). First note that, from (3):

$$\bar{e}_c = E(e_i|c_i) = \gamma_0 + \gamma_1 c_i \tag{5}$$

and:

$$v_i = e_i - (\gamma_0 + \gamma_1 c_i) = e_i - \bar{e}_c \tag{6}$$

Equation (2) can then be rewritten as:

$$h_i = \beta_0 + \beta_1 c_i + \beta_2 \bar{e}_c + \beta_2 v_i + u_i \tag{7}$$

or, in expanded form:

$$h_i = \beta_0 + \beta_1 c_i + \beta_2(\gamma_0 + \gamma_1 c_i) + \beta_2 v_i + u_i = \alpha_0 + \alpha_1 c_i + \beta_2 v_i + u_i \tag{8}$$

In equation (8), the coefficient on c_i gives the total contribution of circumstances and the coefficient on v_i again gives the direct contribution of effort. This provides a connection with the Carrieri and Jones (2018) decomposition which uses variation in c_i to measure the direct contribution of circumstances, variation in $\bar{e}_c$ to measure the indirect contribution of circumstances, and variation in $e_i - \bar{e}_c$ to measure the direct contribution of effort.

In this paper, we use the variance as an inequality measure, given the fact that recent contributions to the IOp literature have favoured the variance as an absolute measure of health inequality (see e.g., Fleurbaey and Schokkaert, 2009, 2012; Jusot, et al., 2013; Carrieri and Jones, 2018). Applying the Shorrocks (1982) decomposition of the variance to equation (7) gives (e.g., Carrieri and Jones, 2018):

$$\mathrm{Var(h)} = \mathrm{cov}(\beta_1 c_i, \mathrm{h}) + \mathrm{cov}(\beta_2 \bar{e}_c, \mathrm{h}) + \mathrm{cov}(\beta_2(e_i - \bar{e}_c), \mathrm{h}) + \mathrm{cov}(u_i, \mathrm{h}) \tag{9}$$

where the first term on the right-hand side relates to the direct contribution of observed circumstances, the second to the indirect contribution of circumstances through effort and the third to the direct contribution of observed effort. In practice $\bar{e}_c$ would be estimated from equation (3).

So far equation (2) has been interpreted as a "parametric" specification that is linear in the observed circumstance and effort variables. In contrast the "nonparametric approach" to IOp creates a dummy variable for each Roemerian type (τ), i.e. for each unique combination of the values of the circumstance variables, and conditions the regression model on these dummy variables rather than the original circumstance variables (e.g., Checchi and Peragine, 2010; Hufe and Peichl, 2015). This is where the curse of dimensionality arises as the number of types can become prohibitively large even with relatively few circumstance variables; for example, with just five circumstance variables, each having five categories, there would be $5^5 = 3{,}125$ unique types. The main advantage of our latent class approach is that it can help to reduce the dimensionality of this problem.

Carrieri and Jones (2018) propose a semiparametric approach in which circumstances are handled nonparametrically by splitting the sample into the separate types (τ) and then using linear regressions of health outcomes on effort variables within those types. To illustrate the connection between their approach and equations (2) and (3) consider the case where there is a single circumstance variable that has only two categories (say, male and female). In this case, the two types correspond to the two values of $c_i \in [0,1]$. Then, using equation (7), the direct contribution of circumstances, $\beta_0 + \beta_1 c_i$, can take two values according to type: β_0 and $\beta_0 + \beta_1$. Similarly, the indirect contribution of circumstances, $\beta_2 \bar{e}_c$, takes two values: $\beta_2 \gamma_0$ and $\beta_2(\gamma_0 + \gamma_1)$; as does the direct contribution of effort, $\beta_2 v_i$: $\beta_2(e_i - \gamma_0)$ and $\beta_2(e_i - \gamma_0 - \gamma_1)$. In fact, the specification proposed by Carrieri and Jones (2018) implies greater flexibility in the direct contribution of effort, such that the slope term is allowed to vary over types. In terms of linear models this implies modifying equation (2) to give:

$$h_i = \beta_0 + \beta_1 c_i + \beta_2 e_i + \beta_3 c_i e_i + u_i \tag{10}$$

Then, for example, the direct contribution of effort takes the two values $\beta_2(e_i - \gamma_0)$ and $(\beta_2 + \beta_3)(e_i - \gamma_0 - \gamma_1)$. The gist of this specification implies a linear relationship between health outcomes and effort in which both the intercept and slopes vary across types.

In the context of the analysis of Carrieri and Jones (2018), this linear model can be extended to a semiparametric specification. Specifically, they partition the sample into distinct types (denoted by τ) for individuals within type τ it is assumed that there is a linear relation between health outcomes and effort:

$$h_i = \theta_{0\tau} + \theta_{1\tau} e_i + u_{i\tau} = \theta_{0\tau} + E_{i\tau} + u_{i\tau} \tag{11}$$

This specification allows inequality indices, such as the Gini coefficient and the variance, to be decomposed. For example, applying the Shorrocks (1982) decomposition of the variance gives the decomposition proposed by Carrieri and Jones (2018):

$$\mathrm{Var(h)} = \mathrm{cov}(\theta_{0\tau} - \bar{\theta}_0, \mathrm{h}) + \mathrm{cov}\big((\bar{\mathrm{E}}_\tau - \bar{\mathrm{E}}), \mathrm{h}\big) + \mathrm{cov}\big((\mathrm{E_i} - \bar{\mathrm{E}}_\tau), \mathrm{h}\big) + \mathrm{cov}(u_{i\tau}, h) \tag{12}$$

The first term in equation (12) is the contribution of the variation of the intercepts $\theta_{0\tau}$ across types, centred at the pooled mean across types. This term measures the direct contribution of circumstances to the overall inequality. The second term reflects the indirect contribution of circumstances to overall inequality, capturing variation in the average level of effort within each type around the pooled mean of effort. The third term is the contribution of the within-type variation in effort to the overall health inequality. In normative terms, this represents the contribution of effort. The final term measures the contribution of residual factors u to overall inequality. Each of these terms are analogous to those derived from the linear model with heterogeneous slopes given by equation (10).

The Carrieri and Jones (2018) specification is fully nonparametric in the way that it handles circumstances, but this comes at the cost of the curse of dimensionality which can limit its applicability. In this paper we propose using a latent class specification to limit the number of types that are modelled, while maintaining the decomposition given by equation (12) based on these latent types. The latent class specification is implemented as a FMM. In the FMM, the conditional density of our health outcome variable, allostatic load, is assumed to be drawn from a population characterised as an additive mixture of K (j=1,…,K) distinct classes in proportions ρ_j, where, $0 \leq \rho_j \leq 1$, $\sum_{j=1}^{K} \rho_j = 1$. Membership of the latent types is unobservable but the mixture probabilities of class membership ρ_j are assumed to be a function of the set of observed circumstance variables ($\boldsymbol{c_i}$):

$$f(h_i|\boldsymbol{c_i}, \boldsymbol{e_i}) = \sum_{j=1}^{K} \rho_j(\boldsymbol{c_i}) f_j\big(h_i\big|\boldsymbol{e_i}, \boldsymbol{\theta_j}\big) \tag{13}$$

$\boldsymbol{\theta_j}$ stands for the vector of parameters describing the conditional density function f_j within each type and $\boldsymbol{e_i}$ is the vector of effort variables.

In FMMs, the prior probability for the j[th] latent class can be expressed as a function of observed circumstance variables using a multinomial logit transformation. For our analysis, we estimate FMMs assuming that the outcome variable (allostatic load) is a mixture of a number of normal distributions, each with its own mean and variance. The normal provides a good fit for our measure of allostatic load. Effectively this fits linear regressions of h_i on e_i for each latent type and the estimates can then be used in the decomposition formula (12). The log-likelihood for the full model is:

$$LogL = \sum_{i=1}^{n} ln\{f(h_i|\boldsymbol{c_i}, \boldsymbol{e_i})\} = \sum_{i=1}^{n} ln\{\sum_{j=1}^{K} \rho_j(\boldsymbol{c_i}) f_j\big(h_i\big|\boldsymbol{e_i}, \boldsymbol{\theta_j}\big)\} \tag{14}$$

For a given value of K, the parameters of the model are estimated jointly by full information maximum likelihood (FIML) using the EM algorithm to refine the starting values.

The choice of the appropriate number of latent types (K) is crucial for FMMs; we use statistical information criteria to identify the FMM with the number of classes that makes the best statistical fit (Cameron and Trivedi, 2010). A caveat for the use of FMMs is the risk that outliers in the data may be captured by additional mixture components. Hence, it is desirable that FMM estimation results in latent classes that account for a sufficient number of observations as well as having meaningful posterior differences in outcomes across the different latent classes (Cameron and Trivedi, 2010; Deb et al., 2011).

Once the number of latent classes (types), K, is selected we can use the parameter estimates from the model to calculate the posterior probability of each individual being assigned to a given latent class $j = 1,2 \ldots K$. The posterior probability of membership in each latent class (type) is calculated conditional on all c, e and the outcome variable (h) as:

$$\Pr(y_i \in type\ \tau | \boldsymbol{\theta}, h_i, c_i, e_i) = \frac{\rho_\tau(c_i) f_\tau(h_i | \boldsymbol{e_i}, \boldsymbol{\theta_v})}{\sum_{j=1}^{K} \rho_j(c_i) f_j(h_i | e_i, \boldsymbol{\theta_j})},\ \forall\ \tau = 1,2, \dots K. \tag{15}$$

For each individual (i), K posterior probabilities are estimated, one for the membership of each type. Following the common practice in the literature (e.g., Deb et al., 2011; Li Donni et al., 2015), we assign each individual to the type with the highest posterior probability (known as *modal assignment*). Given this modal assignment to types we then apply the decomposition analysis based on equation (12) using the parameters estimated for the mixture model[2].

It should be explicitly mentioned here that our paper builds on precedents for the use of latent class models in analysis of health equity (e.g., Bago d'Uva et al., 2009, Balia and Jones, 2011 and Li Donni et al., 2015) to highlight the role that FMMs can play in modelling unobserved heterogeneity (treating Roemerian types as unobserved latent classes). However, we view our approach as a complement to recent data driven perspectives of statistical learning methods (Brunori et al., 2018; Hufe et al., 2019). Specifically, the latter focus mainly on an ex ante IOp approach and the problem of making a parsimonious selection of variables from the observed set of circumstances in a non-arbitrary and data driven way. Our motivation in this paper is to start with a set of observed circumstance variables that accord with the socio-legal context in the UK (as will be justified below) – a normatively driven selection. Then, rather than using the "nonparametric" approach to define types – which results in a prohibitively large number of types – we use latent class specification as a data-driven way of reducing the number of types based on the given set of observed circumstance variables and to address the partial observability of circumstances. It should be stressed here that for a LCM a key choice is which variables to include in the model specification. The degrees of freedom decline not only because of larger number of latent groups but they also decrease with the number of regressors included in the specification. While the set of circumstances considered is a normative issue, it is also a statistical specification issue.

3 Data

The data come from UKHLS, a longitudinal, nationally representative study of the UK. In this study, we use the General Population Sample (GPS) component of

[2] Note that the posterior probabilities use information on all the estimated parameters and all of the data, including the outcomes, to predict membership of the latent types. As a result, our model fit, and associated decomposition analysis explains a much higher proportion of the variation in outcome than using prior probabilities for the modal assignment or using conventional parametric or nonparametric approaches to IOp.

UKHLS, a random sample of the general population. As part of wave 2 (2010-2011), nurse-measured and non-fasted blood-based biomarkers were collected for the GPS.[3] In this study we restricted our sample to adults aged 25 years old and above in order to create more meaningful age groups for the purpose of our analysis to explore differences in IOp in health and its underlying sources across age-sex groups that may reflect different generations; moreover, this allows us to focus on individuals that have completed their educational qualifications, as individual's educational qualifications are included as efforts in our sensitivity analysis.[4] Exclusion of missing data on our biomarkers, circumstance and effort variables results in a working sample of 5,820 adults.

A multi-system biological risk measure: allostatic load

We use a cumulative biomarker index often called allostatic load (e.g., Davillas and Pudney, 2017; Howard and Sparks, 2016; Seeman et al., 2004). The allostatic load is regarded as a biological risk score reflecting the cumulative effects of chronic exposure to psychosocial and environmental challenges or stressors that may leads to significant physiological dysregulation and increased morbidity and mortality risks (Howard and Sparks, 2016; Seeman et al., 2004). As such, allostatic load is of particular relevance in our analysis as IOp is based on concerns about a lasting effect of circumstances on individuals' long-term health.

Our index combines biomarkers for adiposity, blood pressure, inflammation, blood sugar levels and cholesterol (see Table A.1, appendix for a description of the relevant biomarkers). Each of these biomarkers is transformed into standard deviation units and then summed to define allostatic load. It has been shown that a single measure of the different biomarkers is sufficient to measure allostatic load (Howard and Sparks, 2016). Higher values of allostatic load indicate worse health. Given that allostatic load is modelled here as a mixture of normals, it is notable that the density of allostatic load is unimodal and fairly symmetric (Figure 1).

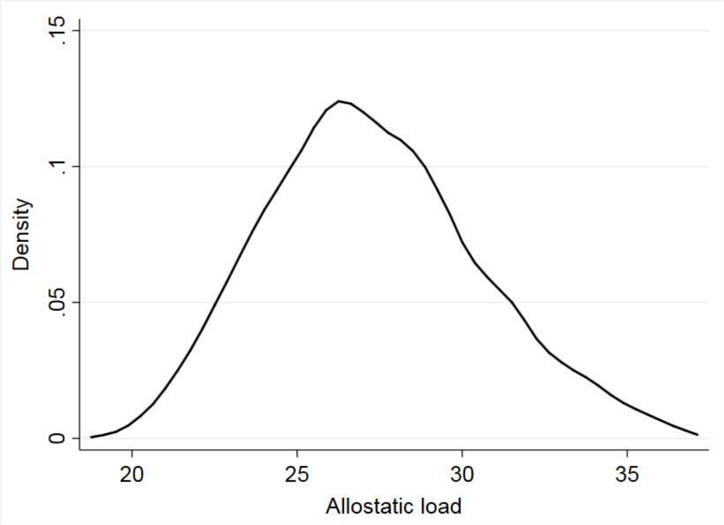
With respect to allostatic load, to illustrate the magnitudes involved, consider a healthy woman with normal waist circumference of 79cm (below the threshold for increased health risks; WHO, 2000) and height 162cm (average height), normal systolic blood pressure (90mmHg; the lower bound for normal blood pressure) and all the other biomarkers (used to define allostatic load) in the population average levels; her allostatic load will be around 24.40. A less healthy woman with a higher waist circumference of 90 cm (above the 88cm cut-point for elevated health risks) and the same body height, high systolic blood pressure 140 mmHg (cut-point for hypertension) and all other biomarkers at their mean values will experience

[3] Respondents were eligible for nurse visits if they were aged 16+, lived in England, Wales, or Scotland, and were not pregnant. Blood sample collections were further restricted to those who had no clotting disorders and no history of fits.

[4] It should be explicitly mentioned here that our results remain practically identical when no age restriction is imposed on our sample.

allostatic load of, 29.10 i.e., a difference of about 4.70 allostatic load points. The range of the allostatic load used in our analysis reflects the individual-level differences in the values of all biomarkers used. Table A1 (Appendix) presents the summary statistics for the allostatic load variable.

Figure 1. Kernel density for the allostatic load index



Circumstances

Our set of circumstance variables embodies the ethical position of the responsibility cut, defining illegitimate sources of health inequality. For the choice of circumstance variables, we follow the recent literature on health equity along with the UK policy and legal context (Davillas and Jones, 2020; Carrieri and Jones, 2018; Rosa Dias, 2009, 2010; Jusot et al., 2013).

Drawing on the socio-legal context in the UK, we treat sex and age as circumstances; sex and age are considered as protected characteristics under the Equality Act of 2010[5]. Specifically, we create three age group dummies (25-44 age group; 45-64 age group; and those 65 and above) for males and females (6 age-sex dummies).[6] Socioeconomic status (SES) in childhood has been an important concern of the existing literature on IOp. Childhood SES is regarded as an important source of IOp in health, being beyond individual's control and exerting a lasting effect on individual's adult health (Jusot et al., 2013; Rosa Dias, 2009, 2010). In our analysis we use both parental occupational status and parental education to proxy childhood SES. Two categorical variables (one for each parent) are used to capture

---

[5] For example, NHS England suggests actions to advance equality of opportunity in health, particularly relevant to patient's age and sex, characteristics that are "protected" under the Equality Act (NHS England, 2017).

[6] These age groups are carefully selected to reflect the three generations: generation X and millennials, the "baby boomers" and the "silent generation". However, the cross-sectional nature of our data does not allow us to make inferences about lifecycle and cohort effects.

the occupational status of the respondent's mother and father, when the respondent was aged 14: not working (reference category), four occupation skill levels and a category for missing data. To construct these variables the occupational skill levels are based on the skill level structure of the Standard Occupational Classification 2010. Given the high correlation between mother's and father's education, we combine them creating a measure capturing the highest parental education level (Kenkel et al., 2006).[7] This is a four-category variable measured as: left school with no/some qualification (reference category), post-school qualification/certificate (e.g., an apprenticeship), degree (university or other higher-education degree) and a missing data category.

Efforts

In the concept of IOp in health, effort variables typically indicate decisions to invest in health capital, such as health-related lifestyles (e.g., Balia and Jones, 2011; Carrieri and Jones, 2018, Rosa Dias, 2010).[8] Smoking status is captured by a categorical variable: current smoker, ex-smoker and never-smoker (reference category). Unhealthy dietary habits are captured by a dummy taking the value of one when the individual does not comply with the recommendation of five portions of fruits or vegetables per day and zero otherwise and an indicator for usual consumption of white (versus non-white) bread. Physical inactivity is captured by a dummy for not being a frequent walker (walk less than 5 times per week) and by a categorical variable for the frequency of sports participation: 3 and more times/week (reference category), 1-3 times/week, once per month or not at all.[9]

---

[7] Due to the high correlation between mother's and father's education we use a combined measure of parental education to alleviate multicollinearity concerns relevant to the multinomial logit model for (latent) class membership component of our FMM. Existing research shows that individuals match on length and type of education, with education being one of the most important mechanisms of homogamy (e.g., Nielsen and Svarer, 2009) and, thus, the observed high correlation between mother's and father's education in our sample is not a new finding in the literature. However, we re-estimated our models to test the robustness of our results in the case that the categorical mother and father education are used separately. The latent type probabilities and decomposition are practically identical to those presented in the paper.

[8] As with most of the equity in health literature, our study does not aim to address causality but to quantify and decompose IOp in health. However, we should highlight that our biological health outcome (allostatic load) captures long-run, systematic exposures to harmful situations rather than contemporaneous effects. As such, concerns regarding reverence causality may be somewhat alleviated as high allostatic load is not directly related to diagnosis of certain current conditions, which may be relevant to GP recommendations for adopting a healthier lifestyle. In any case, accounting for reverse causality is beyond the scope of our study and our results should be interpreted with this in mind.

[9] We experimented with further augmenting our set of effort variables using a binge drinking variable. Specifically, we have used data on the number of pints of beer, spirits, glasses of wine and alcopop that people drank and its frequency (as per number of days per week). We have then transformed these data to alcohol units and created a binge drinking dummy (8 units of alcohol for men and 6 units of alcohol for women on a frequent basis).

For the main analysis, we restrict our effort variables to lifestyle indicators in order to provide clear results on what is the direct contribution of lifestyle in shaping health inequalities and which part of their contribution may operate indirectly via the influence of circumstances (such as family socio-economic status etc.) on individual's lifestyle. Broadly, this is consistent with the existing health economics research that explored the role of lifestyle as a potential mediating factor on the association between socioeconomic status and health (e.g., Balia and Jones, 2008; Baum and Ruhm, 2009). Similar effort variables are used by many of the existing studies in the IOp in health literature (eg., Balia and Jones, 2011, Carrieri and Jones, 2018, Jusot et al., 2013). As stated by Jusot et al. (2013), "Lifestyles, such as doing exercise, having a balanced diet, not smoking or not drinking too much, are widely accepted as examples of efforts in relation to health, representing non-constrained individual choices". However, as a sensitivity analysis, we augment our set of effort variables beyond lifestyle, including individuals' own education, household income and marital status.[10]

4 Results

Table 1 presents the values of the AIC and BIC for each FMM estimated with different numbers of types.[11] The model with three latent classes is the one that minimises the BIC as well as has lower AIC compared to the FMM model with either two or four latent types; thus, selected as our preferable model here (Cameron and Trivedi, 2010). Although the FMM with five latent classes have lower AIC, the BIC value is higher compared to the FMM with three latent classes. Following Cameron and Trivedi (2010), further support for the FMM with three classes comes from the fact that it results in reliably differentiated latent classes (Table 2); each latent class accounts for a sufficient number of observations and the

---

Our FMM results shows that this variable has no systematic effects (at the 10% level) and the decomposition analysis shows that it makes a trivial contribution to the direct role of efforts on explaining total inequality in allostatic load (subsections 4.2 and 4.3). Moreover, the inclusion of this variable reduces our sample size (due to data availability) by about 15%. For these reasons, we decided not to include this variable in the model specification presented in the paper.

[10] For consistency with the other effort variables presented above, these additional variables are coded to reflect a positive association with ill health. Specifically, a categorical variable is used to capture individuals' own education: degree (reference category), A-level and equivalent, GCSE and equivalent and no qualification. A dummy variable for not being married or cohabiting is also included in our augmented effort variables set. Equivalised (using the modified OECD scale) and log transformed household income from all sources is also included; for consistency with all other effort variables, the variable is rescaled to create an index that is decreasing in income.

[11] Table A2 (Appendix) shows the corresponding full set of posterior probabilities and mean allostatic load values by latent class.

mean values of allostatic load (our outcome variables) are distinct across the three latent classes (there is no overlap in their confidence intervals). Table A2 (Appendix) shows that the FMMs with a higher number of latent classes (those with four classes and above), have one or more latent classes that account for a fairly small part of the population and are characterised by non-distinctive latent classes with respect to the predicted allostatic load across types. For example, there is a significant overlap in the 95% confidence intervals of the predicted allostatic load between the latent types 3 and 4 as well as between latent types 4 and 5 for the case of the FMM with four and five latent types, respectively.

Table 1. FMMs for allostatic load: AIC and BIC.

Number of latent classes (types)	AIC	BIC
K=1	29676	29743
K=2	28619	28879
K=3	**28416**	**28862**
K=4	28426	29073
K=5	28377	29183

Notes: AIC and BIC values for the FMMs with a different number of latent classes (types).

Focusing on our preferred FMM with three latent types (Table 2), we find that about 19% of our sample is estimated to belong to type 1 (the latent class with the lowest health risk, i.e., with the lowest mean allostatic load value), 44% in type 2 (the latent class with the second-lowest allostatic load) and 37% in type 3 (the type with the highest allostatic load).

Table 2. Latent class (types) probabilities and predicted mean allostatic load: FMM with three latent types.

	Latent class probabilities ρ_j (%)	Predicted mean allostatic load
Type 1	19.43 (16.08; 23.27)	23.70 (23.35; 24.04)
Type 2	43.96 (33.12; 55.42)	26.81 (26.38; 27.24)
Type 3	36.61 (24.53; 50.64)	30.09 (29.33; 30.85)

Note: 95% confidence intervals in parenthesis

4.1 Modal assignment of individuals to the latent types

As described above, for the needs of our IOp decomposition analysis we use the modal assignment method to classify individuals into the type with the highest posterior probability. Given that it has been shown that modal assignments are problematic when, for a substantial number of individuals, the highest and the next-highest posterior probabilities of belonging to two or more different types are particularly close (e.g., Vermunt and Magidson, 2004), this issue need to be explored here.

Table 3 presents the mean values of the posterior probabilities of class membership conditional on modal type assignment. Focusing on those who are classified into type 1 using the modal assignment ("Type 1" column in Table 3), we find that the mean posterior probability of belonging to type 1 ($\Pr(y_i \in type\ 1)$) is around 81%, with 90% of those individuals having posterior probabilities to belong to this type (i.e., $\Pr(y_i \in type\ 1)$) of above 57.5% (as shown by the relevant quantile statistics; Q10, Q50 and Q90). The corresponding mean posterior probabilities of belonging to types 2 (around 17%) and 3 (around 1.7%) are much lower. Similarly, the mean posterior probability of belonging to type 2 is around 72% for those who are assigned to type 2 using the modal assignment (Table 3, column "Type 2"). Modal assignments to type 3 seem sensible also given the very high mean posterior probability for type 3 membership (around 82%; Table 4, column "Type 3").

Overall, these results show that modal assignments across the three types are sensible in our analysis. For the vast majority of individuals, there are clear differences between the highest posterior probability of belonging to a certain latent type and the other two posterior probabilities for the remaining types.

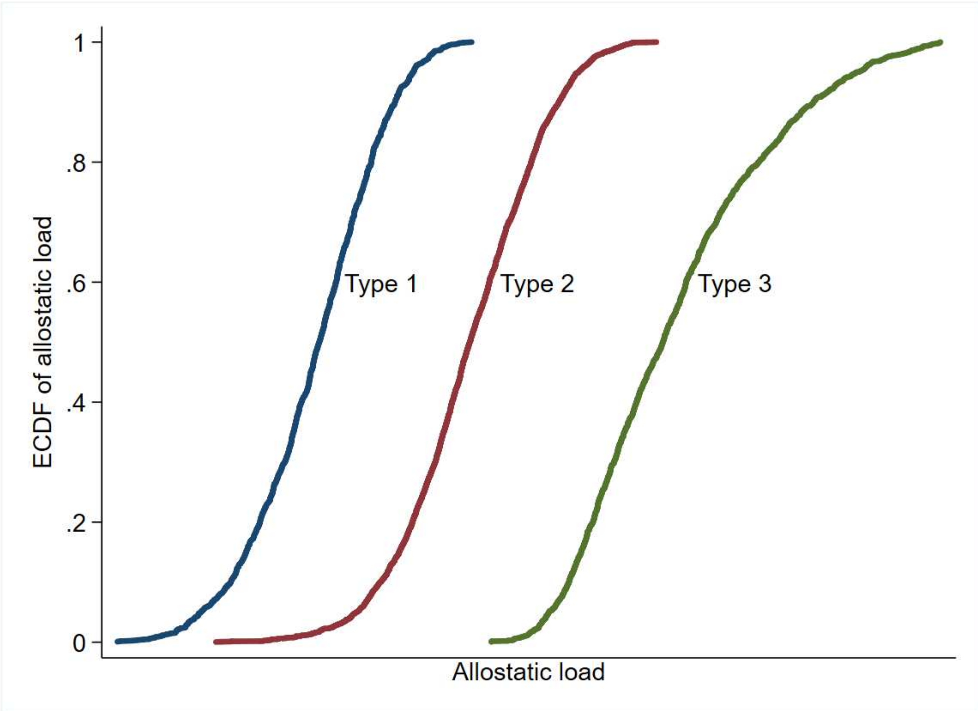
Table 3. Posterior latent class (type) membership probabilities conditional on modal assignment of individuals into types: FMM with three latent types.

	Modal assignment into latent classes (types)		
Posterior probabilities	Type 1	Type 2	Type 3
$\mathbf{Pr}\ (\boldsymbol{y_i \in type\ 1})$			
Mean (i)	**81.4%**	6.6%	0.0%
Q10	57.5%	0.0%	0.0%
Q90	98.4%	26.4%	0.1%
$\mathbf{Pr}\ (\boldsymbol{y_i \in type\ 2})$			
Mean (ii)	17.0%	**71.9%**	18.0%
Q10	1.4%	55.5%	1.4%
Q90	38.5%	87.3%	44.3%
$\mathbf{Pr}\ (\boldsymbol{y_i \in type\ 3})$			
Mean (iii)	1.7%	21.5%	**82.0%**
Q10	0.1%	6.5%	55.7%
Q90	4.1%	41.0%	99.9%
Total (sum of rows i, ii, iii)	100%	100%	100%

Notes: Q10 and Q90 stand for the 10^{th} and 90^{th} quantiles of the posterior probabilities conditional on modal assignment of individuals into types.

Figure 2 presents the graphical illustration of the empirical distribution functions for allostatic load by types, defined using the modal assignment. The graph shows a clear difference in the distribution of allostatic load across types confirming our results in Table 2. From an IOp perspective, these distributions can be interpreted as representing the opportunity sets facing each of the types, in terms of the distribution of health outcomes available to them, bearing in mind that a higher score of allostatic load implies worse health risks; it appears to be first order stochastic dominance across the three types. The contrast between the distributions for types 1, 2 and 3 is striking, with the non-overlapping support for the three distributions suggesting zero order stochastic dominance. As stochastic dominance analysis is often used to test for the presence of IOp, Figure 2 highlights the existence of systematic IOp in allostatic load. This further motivates our analysis below on quantifying and decomposing IOp in allostatic load.

Figure 2. Allostatic load distributions by types (defined using the modal assignment): FMM with three latent types.



4.2. Characterising the profile of the latent types

The analysis so far does not characterise the profile of the three latent types in terms of the observed circumstances. In the concept of IOp, types are defined on the basis of individuals' exposure to circumstance variables and, thus, identifying whether each latent type reflects more or less disadvantaged observed circumstances is of particular importance. Table 4 shows, in each row, the mean posterior probabilities (along with the relevant 95% confidence intervals) of belonging to each of the three latent types ($\Pr(y_i \in type\ 1, 2\ and\ 3)$) conditional on selected observed circumstances; these mean values (and the relevant confidence intervals) are estimated averaging the individual posterior probabilities for the sub-samples of individuals defined based on the observed circumstances (for example, having a mother with highest occupational skill level, etc.). Since, by construction of the latent class model of type membership, each individual is assumed to belong to a single type, the probabilities in each row always add up to 1. For the categorical circumstance variables, mean posterior probabilities are calculated for the most and least deprived category. It should be noted that our set of circumstance variables are jointly highly significant as determinants of individuals' class membership in the multinomial logit model for class membership of our FMM.

Table 4. Posterior type membership probabilities, conditional on observed circumstances: FMM with three latent types.

Observed circumstances	*Type 1*	*Type 2*	*Type 3*
Mother's occupational status			
Highest group (skill level 4)	0.407 [0.378; 0.437]	0.375 [0.337; 0.413]	0.218 [0.192; 0.244]
Lowest group (unemployed)	0.132 [0.121; 0.144]	0.460 [0.448;0.472]	0.407 [0.393; 0.421]
Father's occupational status			
Highest group (skill level 4)	0.302 [0.276; 0.327]	0.464 [0.442; 0.485]	0.234 [0.215; 0.254]
Lowest group (unemployed)	0.169 [0.132; 0.207]	0.503 [0.463; 0.543]	0.328 [0.285; 0.371]
Parental education			
Degree	0.368 [0.333; 0.402]	0.464 [0.435; 0.492]	0.168 [0.146; 0.190]
No qualification	0.145 [0.135; 0.155]	0.449 [0.438; 0.459]	0.406 [0.394; 0.419]
Age-sex profile			
Males 25-44	0.264 [0.241; 0.287]	0.513 [0.492; 0.534]	0.223 [0.201; 0.245]
Males 45-64	0.000 [0.000; 0.000]	0.583 [0.565; 0.602]	0.417 [0.398; 0.435]
Males 65+	0.000 [0.000; 0.000]	0.404 [0.383; 0.425]	0.596 [0.575; 0.617]
Females 25-44	0.604 [0.580; 0.628]	0.260 [0.244; 0.276]	0.136 [0.120; 0.151]
Females 45-64	0.251 [0.234; 0.267]	0.430 [0.417; 0.445]	0.319 [0.301; 0.336]
Females 65+	0.001 [0.001; 0.001]	0.452 [0.431; 0.473]	0.547 [0.526; 0.568]

Notes: The mean values (and the relevant confidence intervals) presented in the table are estimated averaging the individual posterior probabilities for the sub-samples of individuals defined based on the observed circumstances (for example, having a mother with highest occupational skill level, etc.). The probabilities in each row add up to 1. 95% confidence intervals are in brackets.

Younger females, those having a mother (and to lesser extent a father) with higher occupational status as well as those with more educated parents are most likely to belong to type 1 (Table 4). For example, the posterior probability to belong to type 1 for an individual who experienced the most advantaged maternal occupational status during childhood is higher (i.e., 0.407) as compared to type 2 (0.375) and type 3 (0.218). The type 2 latent class lies above the least deprived type (type 1) but is not be considered at the most deprived type among the three. Specifically, although it is more likely to consist of individuals at earlier to later middle ages, those who had a father working in a highly skilled job (skill level 4) and/or at least one parent with a degree qualification, we also observe large posterior probabilities for those at the lowest parental occupation and educational groups to belong to type 2. Type 3 clearly differs from the other two types to the extent that members are more likely to come from those who are older and less likely from those who experienced the higher parental occupation and parental educational status during their childhood. For example, the posterior probability for belonging to type 3 for an individual who experienced the higher parental education (degree) is lower (i.e., 0.168) compared to type 1 (0.368) and type 2 (0.464). Overall, these results reveal a set of three fully characterised latent types, each of which reflects a complex set of observed circumstances. This complex profile of types, obtained using latent class techniques, indicates what may have been missed if single circumstances were chosen to define types.

4.3 Decomposition of overall inequality

The analysis so far shows that modal assignments of individuals into the three latent types are feasible (subsections 4.1 and 4.2). Beyond the definition of types, the FMM analysis also allows us to account for the type-specific heterogeneity in the association between effort variables and allostatic load. Both the latter and the definition of types are of particular importance in our decomposition of inequality analysis.

Specifically, our FMM results show considerable heterogeneity in the association between effort variables and allostatic load (Table A.3, Appendix). Overall, all variables reflecting less healthy lifestyles (given the reference categories) show a positive association with higher allostatic load values indicating higher health risks; the associations become more evident in types 2 and 3, which are the types reflecting more adverse circumstances compared to type 1. A formal statistical test rejects the null hypothesis that the effort coefficients are equal across types (p-value=0.000).

Table 5 presents the results of the decomposition analysis, allowing us to decompose the sources of inequality in allostatic load and the role of IOp on shaping these inequalities based on the results from our FMMs. The table shows the direct contribution of circumstances, the contribution of efforts as well as the indirect contribution of circumstances via efforts to the overall inequality in allostatic load.

We find that the latent types account for most of the total inequality, with the direct contribution of circumstances being the most important component. Specifically, about 50% of the total inequality in our composite health measure is attributed to the direct contribution of circumstances. The contribution of the role of indirect circumstances via efforts show that circumstances exert an indirect contribution to the total inequalities of around 13%, though differences in the association between our effort variables and allostatic load across types. The detailed decomposition of indirect circumstances show that contributions are positive and indicates that the association between the lifestyle variables and allostatic load is larger for the types who have worse health. Lack of frequent physical activity, unhealthy food habits and smoking are the first, second and third most important indirect mechanisms, respectively. Less important however is the direct contribution of the effort variables (within types) in explaining total inequality in allostatic load (accounting for around only 3%)[12].

Table 5. Decomposition of variance in allostatic load based on the FMM with three latent types.

	Absolute contribution†	% contribution†
Direct circumstances	**5.00** [4.87; 5.14]	**49.16 %**
Indirect circumstances via efforts		
Smoking†	0.22	2.16%
Non-compliance: 5 fruits/vegetables	0.25	2.47%
White bread	0.10	1.00%
Non-frequent walking†	0.39	3.83%
Sports activity†	0.39	3.80%
Total indirect circumstances via efforts	**1.35** [1.31; 1.40]	**13.27%**
Direct efforts		
Smoking†	0.10	0.96%
Non-compliance: 5 fruits/vegetables/day	0.01	0.10%
White bread	0.04	0.39%
Non-frequent walking†	0.03	0.29%
Sports activity†	0.15	1.47%
Total direct efforts	**0.32** [0.27; 0.38]	**3.24%**
Residual	**3.50**	**34.50%**
Total Variance	**10.17** [9.84; 10.50]	**100%**

†Absolute and percentage contributions represent the total contribution of all the categories of the relevant categorical variables included in our models.
The decomposition method is described in detail in section 2. The 95% confidence intervals (in brackets) for the direct contribution of circumstances, the total direct contribution of efforts as well as the total indirect role of circumstances via efforts are calculated using a bootstrap with 500 replications.

[12] The decomposition presented in Table 5 relies on modal assignment to latent types based on posterior probabilities. As these probabilities take account of observed outcomes the unexplained variation is much smaller than conventional decomposition methods, with the residual contribution only accounting for 34.5% of the total. For comparison we conducted a decomposition analysis based on the estimated prior probabilities from the FMM, which vary only with observed circumstance variables, and we conducted a decomposition analysis for a standard linear model (see equation 9). The unexplained contribution is much greater for both of these: , 81%for the linear model and 71% for the FMM-prior specification. This larger unexplained component accords with existing decomposition analysis that uses conventional "nonparametric" approaches to define types based on the unique combination of the values of the circumstance variables (Carrieri and Jones, 2018). However, the direct contribution of effort variables remains stable across all three approaches: 3.5% for linear, 3.8% for FMM-prior and 3.2% for FMM-posterior (our main results presented in Table 5).

As discussed above, we conduct a sensitivity analysis by augmenting our set of effort variables beyond lifestyle, including individuals' own education, household income and marital status (Table A.5, Appendix). Augmenting our set of effort variables we find a small reduction in the direct role of circumstances compared to our base case results in Table 5, along with an increase in the indirect role of circumstances via efforts. However, these differences are not large enough to change the conclusions of our analysis in any substantial way; as in the case of our base case results (Table 5), the total (direct and indirect) role of circumstances is still around two thirds of the total inequality in allostatic load (Table A.5, Appendix).

4.4 Decomposition of inequality by age-sex groups

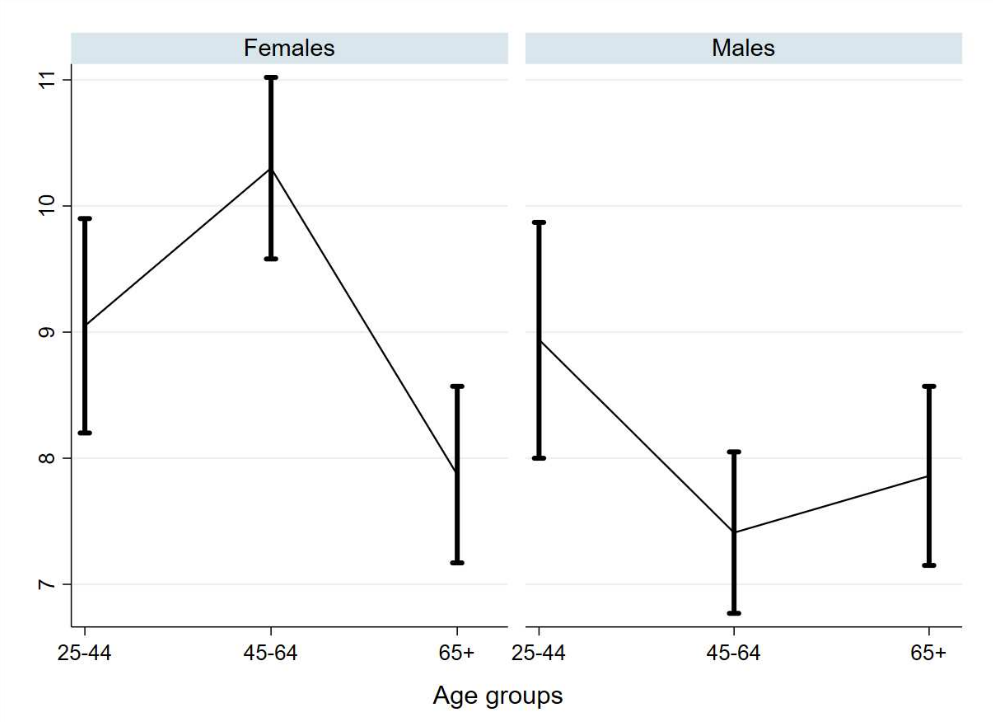
One may argue that although age and sex are not under individual's control, the age-sex variations in allostatic load may be regarded (at least to some extent) as natural or biological and, thus, not seen as a source of unfairness. A way to explore the role of all other circumstance variables, apart from age and sex, to IOp is to undertake post-estimation decomposition analysis in inequalities in allostatic load conditional on the different age-sex groups. Specifically, we use the parameter estimates from our preferred FMM estimated for the full sample and, then, apply the decomposition analysis separately for each of the six mutually exclusive age-sex sub-samples, which are based on the age-sex dummies used in our regression analysis (Data subsection). This allows us to hold sex and age group constant and explore the contribution of all other circumstances.[13] The results of these analyses for the six age-sex groups are presented in this subsection.

Figure 3 shows heterogeneity in the magnitude of total inequalities in allostatic load for our selected age groups. Overall, inequalities are lower for the two older groups of men, while for women an inverse U-shaped pattern is observed. The presence of narrower inequalities at older age groups may be broadly in accordance with the age-as-level hypothesis, which is frequently cited in the health inequalities literature as a hypothesis explaining inequality patterns across the lifecycle (Baum and Ruhm, 2009; Davillas et al, 2017; Davillas and Jones,2020).

---

[13] This analysis by age-sex groups allows us: a) to explore the role of circumstances other than age group and sex; and b) to estimate the coefficients of the other circumstance variables (parental socioeconomic background) after accounting for the potentially confounding role of age group and sex on how parental socioeconomic status may affects individuals' adult health (Baum and Ruhm, 2009). Given that is unlikely to achieve a universal agreement about the treatment of age and sex in the IOp in health literature, our split sample decomposition analysis can be used, supplementary to the full sample decomposition results, to explore the role of our circumstances, apart from age group and sex, in the IOp in health context.

Figure 3. Variance of allostatic load across age groups by sex.



Tables 6 and 7 present the results of the inequality decomposition analysis for the different age groups by sex. For women (Table 6), as in the case of the full sample (Table 5), we find that a substantial part of the total inequalities is attributed to the direct contribution of circumstances across all the age groups. Specifically, the total contribution of circumstances ranges between 40% and 50% across the three age groups, with the direct contribution of circumstances being lower in magnitude in the case of our older age group (accounting for about 40% of the total inequalities). The indirect contribution of circumstances via efforts accounts for between 7% and 17% of the total inequalities, with a higher contribution among younger women. The detailed decomposition of the indirect mechanisms echoes the decomposition results of the full sample with lack of frequent physical activity, unhealthy food habits and smoking being the first, second and third most important contributors, respectively.[14] In line with our results from the pooled

[14] Sports activity (but not the walking variable) has a small but negative contribution in the case of the oldest age group. This may be due to the small sample size in the case of our older age-sex group, which is further split given the three types used in our analysis, as well as because older people are less engaged into sports activities in general. Excluding the sports activity variable from our analysis does not change the overall decomposition results in Tables 6 and 7.

analysis, the direct contribution of efforts is small, ranging from 1.4% to 5.2% across the different age groups.

Table 6. Decomposition of variance in allostatic load by age group: Females

	Age group: 25-44		Age group: 45-64		Age group: 65+	
	Absolute contribution†	% contribution†	Absolute contribution†	% contribution†	Absolute contribution†	% contribution†
Direct circumstances	**4.33 [3.89; 4.77]**	**47.85%**	**5.19 [4.90; 5.49]**	**50.39%**	**3.14 [2.98; 3.31]**	**39.90%**
Indirect circumstances via efforts						
Smoking†	0.33	3.65%	0.25	2.43%	0.05	0.64%
Non-compliance: 5 fruits/vegetables	0.25	2.76%	0.24	2.33%	0.12	1.52%
White bread	0.07	0.77%	0.09	0.87%	0.09	1.14%
Non-frequent walking†	0.20	2.21%	0.37	3.59%	0.45	5.70 %
Sports activity†	0.69	7.62%	0.47	4.56%	-0.10	-1.30 %
Total indirect circumstances via efforts	**1.55 [1.42; 1.67]**	**17.13%**	**1.43 [1.34; 1.52]**	**13.88%**	**0.52 [0.49; 0.55]**	**6.61%**
Direct efforts						
Smoking†	0.06	0.66%	0.08	0.78%	0.11	1.40%
Non-compliance: 5 fruits/vegetables/day	0.00	0.00%	0.01	0.10%	0.02	0.25%
White bread	0.02	0.22%	0.05	0.49%	0.04	0.51%
Non-frequent walking†	0.00	0.00%	0.04	0.39%	0.09	1.14%
Sports activity†	0.04	0.44%	0.16	1.55%	0.15	1.91%
Total direct efforts	**0.13 [0.01; 0.25]**	**1.44%**	**0.34 [0.22; 0.46]**	**3.30%**	**0.41 [0.25; 0.56]**	**5.21%**
Residual	**3.04**	**33.59%**	**3.34**	**32.43%**	**3.83**	**48.67%**
Total Variance	**9.05 [8.20; 9.90]**	**100.00%**	**10.30 [9.58; 11.02]**	**100.00%**	**7.87 [7.17; 8.57]**	**100.00%**
Sample size	**990**		**1,408**		**848**	

†Absolute and percentage contributions represent the total contribution of all the categories of the relevant categorical variables included in our models.
The decomposition method is described in detail in subsection 2.
The 95% confidence intervals (in brackets) for the direct contribution of circumstances, the total direct contribution of efforts as well as the total indirect role of circumstances via efforts are calculated using a bootstrap with 500 replications.

Table 7 presents the decomposition results for men. The direct contribution of circumstances accounts for a substantial part of the total inequalities across all the age groups (ranging from 39% to 47%). The indirect contribution of circumstances via efforts has the second highest contribution to the explained inequalities in allostatic load ranging between 7% and 13%. Once again the direct contribution of efforts accounts for a small part of the total inequalities (ranging between 3% and 6% across the different age groups).

Table 7. Decomposition of variance in allostatic load by age group: Males

	Age group: 25-44		Age group: 45-64		Age group: 65+	
	Absolute contribution†	% contribution†	Absolute contribution†	% contribution†	Absolute contribution†	% contribution†
Direct circumstances	**4.17 [3.75; 4.59]**	**46.64%**	**2.83 [2.66; 3.02]**	**38.19%**	**3.06 [2.89; 3.23]**	**38.93%**
Indirect circumstances via efforts						
Smoking†	**0.25**	2.80%	0.07	0.94%	0.10	1.27%
Non-compliance: 5 fruits/vegetables	**0.24**	2.68%	0.12	1.62%	0.13	1.65%
White bread	**0.07**	0.78%	0.09	1.21%	0.13	1.65%
Non-frequent walking†	**0.25**	2.80%	0.37	4.99%	0.36	4.58%
Sports activity†	**0.30**	3.36%	-0.10	-1.35%	-0.10	-1.27%
Total indirect circumstances via efforts	**1.11 [1.00; 1.22]**	**12.42%**	**0.53 [0.50; 0.56]**	**7.15%**	**0.62 [0.59; 0.65]**	**7.89%**
Direct efforts						
Smoking†	0.09	1.01%	0.14	1.89%	0.09	1.15%
Non-compliance: 5 fruits/vegetables/day	0.00	0.00%	0.01	0.13%	0.02	0.25%
White bread	0.02	0.22%	0.05	0.67%	0.04	0.51%
Non-frequent walking†	0.00	0.00%	0.04	0.54%	0.05	0.64%
Sports activity†	0.17	1.90%	0.23	3.10%	0.16	2.04%
Total direct efforts	**0.29 [0.15; 0.43]**	**3.24%**	**0.45 [0.34; 0.57]**	**6.07%**	**0.35 [0.20; 0.48]**	**4.45%**
Residual	**3.37**	**37.70%**	**3.60**	**48.58%**	**3.83**	**48.73%**
Total Variance	**8.94 [8.00; 9.87]**	**100.00%**	**7.41 [6.77; 8.05]**	**100.00%**	**7.86 [7.15; 8.57]**	**100.00%**
Sample size	**678**		**1,101**		**795**	

†Absolute and percentage contributions represent the total contribution of all the categories of the relevant categorical variables included in our models.
The decomposition method is described in detail in subsection 2.
The 95% confidence intervals (in brackets) for the direct contribution of circumstances, the total direct contribution of efforts as well as the total indirect role of circumstances via efforts are calculated using a bootstrap with 500 replications.

5 Conclusion

A key empirical and practical challenge in all IOp studies is the definition of types. In this paper, we have employed an empirical approach to both analyse and decompose IOp in a composite biomarker measure, allostatic load. Our analysis addresses some of the limitations that affect earlier work, namely the partial observability, the *ad hoc* selection of circumstances and the curse of dimensionality. We use FMMs, a semi-parametric approach to model unobserved heterogeneity regarding type membership, which avoids *a-priori* grouping of individuals into types. This analysis facilitates selection of the number of latent classes (types) and allows us to characterise the latent types in terms of the combination of observed circumstances that they represent, as well as classifying individuals into the different latent types.

For this study we use nationally representative data from the UKHLS. We combine a rich set of nurse-measured and non-fasted blood-based biomarkers to build a cumulative risk score index (also known as allostatic load) which takes into account the chronic exposure to psychosocial and environmental challenges. This allows us to assess the lasting contribution of circumstances and efforts to inequality in long-term health measures.

Our results show a clear ordering of types with respect to both our composite biomarker measure (allostatic load) and the underlying observed circumstances. Beyond the definition of types, FMM analysis allows us to explore the type-specific unobserved heterogeneity in the association between our health measure and efforts, which is crucial for the measurement of *ex post* IOp. Taking advantage of this along with our latent class approach to define types, we have combined the latent class analysis with a recently developed decomposition technique on IOp in health (Carrieri and Jones, 2018). Our more parsimonious and data-driven definition of types (using a latent class model framework) are of importance given the recent evidence that a large number of types may create upward bias in the IOp measurement (Brunori et al., 2019).

We find that a latent class model with three unobserved types provides the best fit with our data, indicating that a relatively small number of types are enough to characterise the sample. Our results show that the characteristics of each of these types reflect a complex combination of observed circumstances, which may be missed if single circumstances or *ad hoc* selections of circumstances were chosen to define types. After classifying individuals into the latent types using modal assignments, we decompose overall inequality in allostatic load. We find that the sum of all sources of inequality in allostatic load attributable to these types (direct effect of circumstances and indirect via their influence on efforts) is about 63% (about 50% due to direct role of circumstances). On the other hand, legitimate sources of inequality (the direct contribution of efforts), which are consistent with the reward principle, account for only around 3% of the total inequality. Further, postestimation inequality decomposition analysis conditional on our selected age-sex groups reveal that, although total inequalities in allostatic load vary across the adult age span, the main conclusions of our study remained mostly unaffected; the relative (percentage) contribution of all other circumstance variables (both direct and indirect) still account for around two thirds of the total inequalities in allostatic load across the different age-sex groups and the direct contribution of observed effort is small.

References

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

Bago d'Uva, T. and Jones, A.M. (2009) "Health care utilisation in Europe: new evidence from the ECHP", *Journal of Health Economics* 28, 265-279.

Bago d'Uva, T., Jones, A.M. and van Doorslaer, E. (2009) "Measurement of horizontal inequity in health care utilisation using European panel data", *Journal of Health Economics*, 28, 280-289.

Bago d'Uva, T., O'Donnell, O., Van Doorslaer, E. (2008). Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. *International Journal of Epidemiology*, 37, 1375-1383.

Balia, S., Jones, A.M. (2011). Catching the habit: a study of inequality of opportunity in smoking-related mortality. *Journal of the Royal Statistical Society Series A*, *174*, 175-194.

Baum II, C.L., Ruhm, C.J. (2009). Age, socioeconomic status and obesity growth. *Journal of Health Economics*, 28(3), 635-648.

Bond, T.N., Lang, K. (2019). The sad truth about happiness scales. *Journal of Political Economy*, 127, 1629-1640.

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

Brunori, P., Hufe, P., Gerszon Mahler, D. (2018). The roots of inequality: estimating inequality of opportunity from regression trees. Policy Research Working Paper 8349, *World Bank Group*, Washington DC.

Brunori, P., Peragine, V., Serlenga, L. (2019). Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare*, 52: 635-661.

Cameron, A.C. and Trivedi, P.K., 2010. Microeconometrics Using Stata. College Station, TX: Stata press.

Carrieri, V., Jones, A.M. (2018). Inequality of opportunity in health: A decomposition-based approach. *Health Economics*, 27, 1981-1995.

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

Crossley, T. F., Kennedy, S. (2002). The reliability of self-assessed health status. *Journal of Health Economics*, 21, 643-658.

Davillas, A., Benzeval, M., & Kumari, M. (2017). Socio-economic inequalities in C-reactive protein and fibrinogen across the adult age span: Findings from Understanding Society. *Scientific reports*, *7*(1), 2641.

Davillas, A., Jones, A.M. (2020). Ex ante inequality of opportunity in health, decomposition and distributional analysis of biomarkers. *Journal of Health Economics*, 69.

Davillas, A., Pudney, S. (2017). Concordance of health states in couples: analysis of self-reported, nurse administered and blood-based biomarker data in the UK understanding society panel. *Journal of Health Economics*, 56, 87-102.

Deb, P., Gallo, W.T., Ayyagari, P., Fletcher, J.M., and Sindelar, J.L. (2011). The effect of job loss on overweight and drinking. *Journal of Health Economics*, 30, 317-327.

Deutsch, J., Alperin, M.N.P., Silber, J. (2018). Using the Shapley Decomposition to Disentangle the Impact of Circumstances and Efforts on Health Inequality. *Social Indicators Research*, 138, 523-543.

Ferreira, F.H., Gignoux, J. (2013). The measurement of educational inequality: Achievement and opportunity. *The World Bank Economic Review*, 28, 210-246.

Fleurbaey, M., Peragine, V. (2013). Ex ante versus ex post equality of opportunity. *Economica*, 80, 118-130.

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, 1348-1367.

Howard, J. T., and Sparks, P. J. (2016). The effects of allostatic load on racial/ethnic mortality differences in the United States. *Population Research and Policy Review*, 35(4), 421-443.

Hufe, P., Peichl, A. (2015). Lower bounds and the linearity assumption in parametric estimates of inequality of opportunity. *IZA Discussion Paper No. 9605.*

Hufe, P., Peichl, A., Weishaar, D. (2019). Lower and upper bounds of inequality of opportunity in emerging economies. Ifo Working Paper Seroes 301, ifo Institute – Leibniz Institute for Economic Research at the University of Munich.

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

Kenkel, D., Lillard, D., Mathios, A. (2006). The roles of high school completion and GED receipt in smoking and obesity. *Journal of Labor Economics*, 24, 635-660.

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.

Lefranc, A., Trannoy, A. (2017). Equality of opportunity, moral hazard and the timing of luck. *Social Choice and Welfare*, 49(3-4), 469-497.

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.

McEwen, B.S., (2015). Biomarkers for assessing population and individual health and disease related to stress and adaptation. *Metabolism*, 64(3), S2-S10.

NHS England (2017). NHS England's response to the specific equality duties of the Equality Act 2010. London: NHS England.

Nielsen, H. S., Svarer, M. (2009). Educational homogamy. How much is opportunities? *Journal of Human Resources*, 44(4), 1066-1086.

Ramos, X., Van de Gaer, D. (2016) Approaches to inequality of opportunity: principles, measures and evidence. *Journal of Economic Surveys*, 30, 855-883.

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

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, 252-264.

Seeman, T.E., Crimmins, E., Huang, M. H., Singer, B., Bucur, A., Gruenewald, T., Berkman L.F., Reuben, D.B. (2004). Cumulative biological risk and socio-economic differences in mortality: MacArthur studies of successful aging. *Social Science & Medicine*, 58, 1985-1997.

Shorrocks, A.F. (1982). Inequality decomposition by factor components. *Econometrica*, 193-211.

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

Vermunt, J.K., Magidson, J. (2004). Latent class analysis. *The Sage Encyclopedia of Social Sciences Research Methods*, 2, 549-553.

WHO (2000). Obesity: preventing and managing the global epidemic (No. 894). *World Health Organization*, Geneva, Switzerland.

Appendix

Table A.1 Description of biomarkers used for allostatic load

Biomarker	Description	Mean	Standard deviation	Lowest value	Highest value
Waist-to-height ratio (WHR)	Waist circumference (cm) over height (cm)	0.563	0.077	0.363	0.94
Systolic blood pressure (SBP)	Maximum pressure in an artery when the heart is pumping blood (mmHg)	126.9	16.371	80.5	209.0
C-reactive protein (CRP)	Inflammatory biomarker; rises as part of the immune response to infection (mg/L)	2.049	1.961	0.20	10.0
Fibrinogen	Fibrinogen (g/L) is a glycoprotein that aids the body to stop bleeding by promoting blood clotting, and is regarded as an inflammatory biomarker.	2.756	0.508	1.10	5.0
Glycated haemoglobin (HbA1c)	Blood sugar biomarker; diagnostic test for diabetes. (mmol/mol)	37.0	6.483	25.0	96.0
Cholesterol ratio	Fat in the blood biomarker; ratio of the total cholesterol (mmol/L) over the high-density lipoprotein cholesterol (mmol/L).	3.739	1.310	1.161	11.14
Allostatic load	Allostatic load is defined as a cumulative measure with each of the biomarkers above transformed into standard deviation units and then summed.	27.422	3.190	19.30	36.62

Table A.2 Latent class probabilities and predicted mean allostatic load: FMMs with different number of latent classes.

Number of latent classes (types)		Latent class probabilities (%)	Predicted mean of allostatic load
K=2	Type 1	26.74 (23.86; 29.84)	24.43 (24.20; 24.65)
	Type 2	73.26 (70.16; 76.14)	28.49 (28.37; 28.62)
K=3	Type 1	19.43 (16.08; 23.27)	23.70 (23.35; 24.04)
	Type 2	43.96 (33.12; 55.42)	26.81 (26.38; 27.24)
	Type 3	36.61 (24.53; 50.64)	30.09 (29.33; 30.85)
K=4	Type 1	19.73 (16.51; 23.41)	23.70 (23.38; 24.04)
	Type 2	41.93 (30.61; 54.16)	26.80 (26.35; 27.25)
	Type 3	5.56 (2.26; 13.03)	29.38 (28.87; 29.89)
	Type 4	32.78 (20.43; 48.08)	30.07 (29.13; 31.00)
K=5	Type 1	18.52 (15.45; 22.06)	23.59 (23.20; 23.97)
	Type 2	5.07 (1.54; 15.41)	26.06 (25.83; 26.29)
	Type 3	43.95 (32.96; 55.57)	27.07 (26.63; 27.51)
	Type 4	4.30 (2.49; 7.33)	30.13 (29.38; 30.87)
	Type 5	28.16 (16.19; 44.30)	30.33 (29.14; 31.52)

Notes: 95% Confidence intervals in parenthesis

Table A.3 Heterogeneous association between efforts and allostatic load by latent type: regression coefficients (and standard errors).

	Type 1	Type 2	Type 3
Current smoker	0.441** (0.171)	0.820*** (0.177)	1.115*** (0.215)
Ex-smoker	-0.008 (0.129)	0.337*** (0.110)	0.407*** (0.154)
Non-compliance:5 fruits/vegetables/day	-0.103 (0.141)	0.094 (0.119)	0.244 (0.170)
White bread	0.284** (0.142)	0.244*** (0.131)	0.543*** (0.159)
Non-frequent walking	0.111 (0.119)	0.063 (0.130)	0.688*** (0.149)
Sports activity: 1-3 times/week	-0.008 (0.170)	0.530*** (0.180)	0.122*** (0.265)
Sports activity: at least once/month	0.106 (0.186)	0.921*** (0.210)	0.520*** (0.303)
Sports activity: less frequent/not at all	0.346** (0.176)	1.142*** (0.158)	1.073*** (0.233)
Constant term	23.387*** (0.200)	25.579*** (0.256)	28.426*** (0.455)

Standard errors in parenthesis.
* p-value<0.1; ** p-value<0.05; *** p-value<0.01.

Table A.4 Multinomial logit specification for (latent) class membership from the FMM model with three latent classes (type 1 base outcome)

	Type 2		Type 3	
	Coeff.	Std. Err.	Coeff.	Std. Err.
Females: Age 45-64	1.356***	0.179	1.661***	0.176
Females: Age 65+	7.011	7.963	7.793	7.961
Males : Age 25-44	1.536***	0.215	1.379***	0.229
Males: Age 45-64	19.866***	0.699	20.156***	0.805
Males: Age 65+	19.131***	2.725	20.118***	2.689
Mother's occupation: skill 1 (lowest)	-0.052	0.244	0.095	0.206
Mother's occupation: skill 2	-0.206	0.162	-0.118	0.150
Mother's occupation: skill 3	0.048	0.237	-0.160	0.233
Mother's occupation: skill 4 (highest)	-0.524**	0.224	-0.666***	0.233
Mother's occupation: missing	-0.009	0.476	0.144	0.443
Father's occupation: skill 1 (lowest)	-0.259	0.414	-0.031	0.376
Father's occupation: skill 2	-0.595	0.358	-0.350	0.329
Father's occupation: skill 3	-0.659**	0.341	-0.472	0.319
Father's occupation: skill 4 (highest)	-0.965***	0.365	-1.073***	0.350
Father's occupation: missing	-0.537	0.416	-0.027	0.364
Parental education: post-school qualification	-0.129	0.162	-0.193	0.147
Parental education: degree	-0.021	0.230	-0.502**	0.253
Parental education: missing	-0.147	0.209	-0.018	0.189

* p-value<0.1; ** p-value<0.05; *** p-value<0.01.

Table A.5 Decomposition of variance in allostatic load: Augmented effort variables.

	Absolute contribution†	% contribution†
Direct circumstances	**4.21** [4.08; 4.31]	**41.39%**
Indirect circumstances via efforts		
Smoking†	0.21	2.06%
Non-compliance: 5 fruits/vegetables	0.18	1.77%
White bread	0.10	0.98%
Non-frequent walking†	0.40	3.93%
Sports activity†	0.38	3.74%
Individual's education†	0.00	0.00%
Household Income†	0.63	6.11%
Married	0.21	2.06%
Total indirect circumstances via efforts	**2.11** [2.051; 2.17]	**20.65%**
Direct efforts		
Smoking†	0.07	0.69%
Non-compliance: 5 fruits/vegetables/day	0.01	0.10%
White bread	0.02	0.20%
Non-frequent walking†	0.04	0.39%
Sports activity†	0.15	1.47%
Individual's education†	0.09	0.88%
Household Income†	0.04	0.39%
Married	0.02	0.20%
Total direct efforts	**0.44** [0.37; 0.50]	**4.33%**
Residual	**3.42**	**33.63%**
Total Variance	**10.17** [9.84; 10.50]	**100%**

†Absolute and percentage contributions represent the total contribution of all the categories of the relevant categorical variables included in our models.

The decomposition method is described in detail in subsection 2. The 95% confidence intervals (in brackets) for the direct contribution of circumstances, the total direct contribution of efforts as well as the total indirect role of circumstances via efforts are calculated using a bootstrap with 500 replications.