INEQUALITY IN HEALTH VERSUS INEQUALITY IN LIFESTYLES

Arnstein Øvrum[1,2] and Kyrre Rickertsen[1,2]

[1]Norwegian Agricultural Economics Research Institute, Oslo, Norway, and
[2]Department of Economics and Resource Management, Norwegian University of Life Sciences, Ås, Norway
e-mail: arnstein.ovrum@nilf.no

Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

INEQUALITY IN HEALTH VERSUS INEQUALITY IN LIFESTYLES

Abstract
This paper uses repeated cross section data from Norway to compare patterns of inequality in self assessed health and obesity with patterns of inequality in underlying lifestyles central to the production of good health, namely physical activity, non-smoking and diet quality, represented by fish and fruits and vegetables consumption. We estimate a multivariate probit model to study correlates of these lifestyle and health variables, while Gini and concentration indices are being decomposed to identify sources of inequality. Results point towards considerable heterogeneity across the different lifestyle and health variables. Thus, patterns of inequality in health outcomes are not necessarily representative of patterns of inequality in their underlying production factors, including lifestyles. While education is generally found to be an important source of inequality, and in some variables, primarily the health variables, also income, there are several cases in which other factors are much more important in explaining inequality, such as gender in fruits and vegetables eating, age in fish eating, and maternal education in obesity. Assuming that poor health and health inequality should ideally be prevented rather than treated, policies should mainly focus on production factors of health, including lifestyles, rather than final health itself. To be efficient, however, the design of such policies may need to be based on lifestyle-specific knowledge, as suggested by our heterogeneous results.

1. Introduction

Observable lifestyle choices such as non-smoking, physical activity and healthful diets are central to the production of good objective health, including the prevention of premature death and chronic disease. Lifestyles are also important predictors of self assessed health (SAH) and obesity, which are in turn closely related to objective health (WHO, 2003). Analogously, lifestyles are expected be an important source of inequality in these health measures, that is, they are expected to be important in explaining why health is unequally distributed across individuals within a population, including in particular individuals of different socioeconomic status.

In recent years, considerable efforts have been made in trying to improve our understanding of patterns of inequalities in health, including attempts at identifying their origins, or sources. In particular, decomposition techniques for the Gini index and the Concentration index (CI) have helped identify the relative contribution of different factors – e.g., of different socio-demographic variables – to total inequality in health and socioeconomic inequality in health, respectively (e.g., van Doorslaer and Jones, 2003). Some of the studies using these decomposition techniques have considered the role of lifestyles in possibly explaining inequality in health and in mediating the direct effect of socioeconomic status on health (Balia and Jones, 2008; Costa-Font and Gil, 2008; Vallejo-Torres and Morris, 2010).[1] However, to the best of our knowledge, no studies have yet made lifestyles themselves be the subject of investigation in such decomposition analyses.

To reduce inequalities in health, and in particular those related to socioeconomic status, is stated as a key goal for health policy in many countries (Marmot et al., 2008). Policies that seek to address health inequalities should ideally be targeted towards production

---

[1] Using British panel data, Balia and Jones (2008) found that lifestyles play an important role in explaining the Gini index for predicted mortality (25.3 % in total), and that, when being appropriately accounted for econometrically, these lifestyles reduce the direct role of socioeconomic status in predicting mortality. Costa-Font and Gil (2008) found that in Spain, their variables for physical activity, smoking and food habits explained respectively 5.8, 2.6 and 0.12 % of the income-related CI for obesity, while Vallejo-Torres and Morris (2010) found that in England, smoking explained 2.3 % of the income-related CI for health, as measured by EQ-5D.

factors of health, including lifestyles, and not final health itself. In order for such policies of 'preventive medicine' to be efficient, however, there is need for more insights into patterns of inequality across several important health affecting lifestyles, including the extent to which these patterns are similar to those of final health. If patterns are homogeneous, then policies that are formulated with the intention of improving health affecting behaviours on the basis of knowledge about inequality in final health are relevant due to their 'trickle down' properties. If on the other hand patterns of inequality are significantly different across different lifestyles and final health, this no longer holds, which if true may have important implications for policy.

The main goal of this paper is therefore to directly compare patterns of inequality in health with patterns of inequality in lifestyles believed to be important in affecting health. Our intention is therefore to add to the existing literature, which has mainly focused on patterns of inequality in final health, where the role of lifestyles, where considered, has been limited to their relative contribution to total inequality or socioeconomic inequality in health, as one out of many possible sources of inequality. Since the empirical evidence generally suggest that lifestyles are important in affecting health, and furthermore, that similar factors, such as socioeconomic status, tend to be important in predicting both health related lifestyles and different health outcomes, we hypothesize that, in one single sample, patterns of total inequality and socioeconomic inequality are similar across various production factors of health, here represented by lifestyle variables, and final health itself.

Our data are drawn from the Norwegian Monitor Survey 2005–2009, which is a repeated cross section survey. Health status, or final health, is proxied by SAH and obesity, while lifestyles are represented by physical activity, smoking, and two indicators of diet quality – frequencies of eating fish and fruits and vegetables.[2] In our analysis, we first study correlates and determinants of lifestyles, obesity and SAH using a multivariate probit model, and next we study sources and patterns of inequality in these variables by decomposing Gini indices and income- and education-related CIs. As explanatory factors, and sources of inequality, we consider (i) a standard set of socio-demographic variables, including current income and education, (ii) a set of variables which represent childhood circumstances, including parental education, and (iii) a set of variables which are meant to serve as proxies for time preferences, risk aversion, and self control. Evidence on the importance of psychological traits (e.g, Heckman, 2007) and childhood circumstances (e.g., Case et al., 2005) in affecting adult lifestyles and health are accumulating. Recent evidence from the US, Britain and France on some of these issues are provided in Cutler and Lleras-Muney (2010), Rosa Dias (2009, 2010) and Trannoy et al. (2010).

2   Data and methods
This study uses data from the Norwegian Monitor Survey, a nationally representative and repeated cross section survey of Norwegian adults which has been conducted biannually since 1985. Some of the key variables that are being analyzed in this paper are based on survey questions that were first introduced in the 2005 survey.[3] Thus, in this paper, only data from the 2005, 2007 and 2009 survey rounds are being used. Individuals aged 15–95 years are recruited to participate in the survey. We restrict our sample to only include respondents aged 25–74 years, as we want to study individuals who can be expected to having completed most of their education and started earning incomes, and since few respondents in the age range

---

[2] In this paper, obesity is defined as a health variable, as being obese is not only an intermediate risk factor for chronic disease and premature death, but also, represents a direct cause of reduced physical and mental health (WHO, 2000, 2003).
[3] These questions relate to self-reported height and body weight, and parental education.

75–95 years are included in the survey. After deleting observations with missing information on any relevant variables (1995 observations), our final sample consists of 7738 observations.

Table 1. Summary statistics

Variable	Description	Mean	S.D.
Lifestyles			
PA	Do physical activity at least twice per week: 1	0.579	0.494
FV	Eat fruits and vegetables at least twice per day: 1	0.387	0.487
FISH	Eat fish for dinner at least once per week: 1	0.765	0.424
NSMOKE	Not smoking cigarettes daily: 1	0.775	0.418
Health			
NOBESE	Body mass index (weight in kg/height in meter^2) < 30: 1	0.876	0.329
SAH	Self assessed health 'good' or 'very good': 1	0.725	0.447
Demographics	*(Ref. categories are "Age 25-34" and "(Living as) married")*		
Age 35-44	Age 35-44: 1	0.262	0.440
Age 45-54	Age 45-54: 1	0.214	0.410
Age 55-64	Age 55-64: 1	0.189	0.392
Age 65-74	Age 65-74: 1	0.129	0.336
Female	Female: 1	0.477	0.500
Household has kids	If any children is living in household: 1	0.494	0.500
Widow	If widowed: 1	0.037	0.188
Divorced	If divorced: 1	0.080	0.272
Single	If single: 1	0.119	0.323
Socioeconomic status	*(Ref. cats. are "Secondary school" and "Income quartile 1")*		
High school	If highest education is high school: 1	0.333	0.471
Some college	If highest education is some university/college: 1	0.195	0.396
College degree	If highest education is university/college with degree: 1	0.360	0.480
Income quartile 2	If household income in 2nd quartile: 1	0.270	0.444
Income quartile 3	If household income in 3rd quartile: 1	0.230	0.421
Income quartile 4	If household income in 4th quartile: 1	0.251	0.434
Occupation	*(Ref. category is "Non-manual worker")*		
Skilled manual	If skilled manual worker: 1	0.196	0.397
Unskilled manual	If unskilled manual worker: 1	0.070	0.256
Social security	If on social security/benefit: 1	0.082	0.275
Other occupation	If unemployed, student, homemaker, retired, or other: 1	0.259	0.438
Psychological traits			
Pay in installments	Like to pay in instalments: 1[a]	0.155	0.362
Life insurance	Household has purchased life insurance: 1	0.479	0.500
Self control	Feel self-control over life outcomes: 1[b]	0.846	0.361
Childhood conditions	*(Ref. cats. are poor childhood and lower parental education)*		
Childhood ec. average	If family's economic situation normal when 10-15 years: 1	0.655	0.475
Childhood ec. rich	If family well-endowed when 10-15 years: 1	0.137	0.344
Mother high school	If mother's highest education level high school: 1	0.209	0.407
Mother college	If mother's highest education level university/college: 1	0.150	0.358
Father high school	If fathers's highest education level high school: 1	0.217	0.412
Father college	If fathers's highest education level university/college: 1	0.222	0.416

Notes: Summary statistics using sample weights. Data pooled from survey years 2005, 2007 and 2009, in total 7738 individual observations. [a]Respondent "partly agrees" or "totally agrees" in that he/she likes to purchase in instalments. [b]Respondent "partly disagrees" or "totally disagrees" in the statement "It is of little use to plan for the future, since what happens in life is mostly a matter of being lucky or unlucky anyway". The original survey question on household income has nine response alternatives, each of which represents a specific income interval. Before dividing income into four classes, we (i) set household income to mid point values of each income interval, (ii) adjusted for inflation over the survey period 2005–2009, and (iii) adjusted for household size by dividing household income by the square root of household size (OECD, 2008).

Each respondent answers an extensive list of questions, of which those related to the selected lifestyle variables and SAH are based on various types of categorical scales. For example, the respondents are asked to indicate their frequency-of-eating (i) fruits and berries and (ii) vegetables on a ten-point scale ranging from *'Never/Less than once per month'* to *'4 times per day'*. Similarly, physical activity has an eight-point frequency-scale ranging from *'Never'* to *'Once or more per day'*. Yet other frequency-scales are being used for smoking and fish consumption. SAH is based on a five-point scale ranging from *'Very bad'* to *'Very good'*. The use of different categorical scales complicates our intention to compare these variables as well as body mass with respect to their determinants and their patterns of inequality. We therefore choose to dichotomize each of these variables, although some information is lost in doing so. Their definition along with summary statistics for other relevant variables of this study is presented in Table 1. The usual disclaimer applies with respect to strengths and limitations of using SAH and obesity as indicators of health, including, amongst others, the possibility of respondents under-reporting their weight and over-reporting their height (Connor Gorber et al., 2007).

Our methodological approach to comparing sources and patterns of inequality in lifestyle and health variables is inspired by related empirical literature, in particular Balia and Jones (2008). The determinants and correlates of lifestyles, obesity and SAH will be studied using a six-equation multivariate probit model, while measures of total inequality and income- and education-related inequality in these variables will be studied using Gini and CI decomposition techniques.

The multivariate probit model of this study includes all the variables listed in Table 1, in addition to controls for survey years. Thus, the dependent variables in this model are our six lifestyle and health variables, while control groups include basic demographic variables, income, education, occupational status, psychological traits, childhood conditions, and survey years.[4] We use identical regressors in all six lifestyle and health equations. Thus, we do not estimate a recursive system in which lifestyles are assumed to affect health, as in for example Balia and Jones (2008). There are two main reasons for this approach. First, our main interest lies in directly comparing important lifestyle and health variables with respect to their determinants and patterns of inequality, rather than in assessing the actual impact that different lifestyles are having on health. Second, unlike in Balia and Jones (2008), our data is not longitudinal, which means that we are only able to assess the impact of current lifestyles on current health, an approach which is mainly relevant for the unknown respondents for whom current lifestyles are good proxies of past lifestyles, as the impact of lifestyles on health is not immediate, but rather, is the result a long-lasting, cumulative process.

As our multivariate probit model for lifestyles and health is not recursive, its main advantage over single equation probit models is its ability to estimate correlation coefficients between error terms of the different equations in the system. Thus, with the multivariate probit model, we can learn about the extent to which unexplained residuals of variation, or unobserved individual characteristics, are systematically related across the different lifestyle and health equations (Balia and Jones, 2008). Technically this is accomplished by utilizing properties of the multivariate normal distribution. The vector or error terms $\boldsymbol{\varepsilon}$ in the multivariate probit model is distributed multivariate standard normal, $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Omega})$, with

---

[4] There potentially exist many and complex interrelations between these different control variables, and also between these control variables and the different dependent variables, including, amongst others, issues related to the direction of casual effects. Furthermore, our data are of cross sectional nature, as noted above. Consequently, we will generally not attempt at making causal inference in this study. Still we claim to be making a novel contribution, by directly comparing important lifestyle and health variables with respect to their correlates and sources of inequality, using one single data set, and a rich set of relevant control variables.

our 6×6 variance-correlation matrix $\boldsymbol{\Omega}$ having values of 1 on its leading diagonal elements, and symmetrical correlation coefficients ρ_{jk} between equations j and k on its off-diagonal elements.

Our multivariate probit model is estimated using simulated maximum likelihood, with the 7738×6 matrix of lifestyle and health probabilities being simulated using the Geweke–Hajivassiliou–Keane (GHK) simulator.[5] More details on the properties and technicalities of the multivariate probit model, including advantages of using the GHK simulator, can be found in Cappellari and Jenkins (2003), Contoyannis and Jones (2004), and Balia and Jones (2008).

The Gini index is a widely used measure of total inequality in a specific variable – it measures the extent to which for example SAH is unequally distributed within a population. The Gini has range [0, 1]. The CI is a closely related measure of socioeconomic-status-related-inequality in, say, the same variable – that is, it measures the extent to which the distribution of SAH is related to a specific measure of socioeconomic status, for example education. The standard version of the CI (Gini) may be expressed as follows;

$$CI_s = \frac{2}{\mu} \, \text{cov}(y, \, r), \tag{1}$$

where r in the case of CI indicates the fractional rank of the chosen socioeconomic indicator, for example education, y is the other variable from which to calculate CI, for example SAH, and μ is the mean of y.[6] The CI has range [-1, 1], where 1 (-1) indicates extreme cases in which all 'good health' is found among those with the absolute highest (lowest) socioeconomic status. The Gini index of total inequality in health is obtained simply by replacing r of socioeconomic status in Eq. (1) by r of health, i.e., by the fractional rank of health (van Doorslaer and Jones, 2003).

Because covariances are central to both the CI and Gini (Eq. 1), these indices may both be obtained, or calculated, using linear regression. An extension of this property is that they may also be decomposed into their contributing factors (Wagstaff et al., 2003). Thus, one might estimate the percentage contribution of e.g. age, gender and education to total inequality in SAH. In this paper, in Section 4, we will decompose both the Gini index of total inequality and the CIs of income- and education-related inequality in our six lifestyle and health variables. The decomposition formula for the CI (Gini) is

$$CI_y = \sum_k \left(\frac{\beta_k \mu_k}{\mu_y} \right) CI_k + \frac{CC_\varepsilon}{\mu_y}, \tag{2}$$

where $(\beta_k \mu_k / \mu_k)$ is the elasticity of variable k, e.g. gender, with respect to y, e.g. SAH, with μ_k and μ_y being the mean of k and y, and with β_k being the coefficient for regressor k in a linear regression on y. CI_k is the CI of variable k with respect to the chosen socioeconomic status indicator, for example education. CC_ε/μ_y is the generalized CI for the error term, which can be computed as a residual (Balia and Jones, 2008). Thus, as an example, in order for gender to make a substantial contribution to the explained part of the education-related CI in SAH, we must have that (i) the elasticity of gender on SAH, controlling for other factors, is large, and (ii) gender and education are strongly correlated, i.e. CI_{gender} is large. The Gini is also

---

[5] As programming the optimization routine for the multivariate probit model is fairly involved, we used the Stata module *mvprobit* for estimation, which was developed by Cappellari and Jenkins (2003). We follow their recommendation of using about $\sqrt{N}$ draws for the GHK simulator (we use 90 draws).

[6] The fractional rank of a variable X is the integer rank of X divided by the sample size N, i.e. $r_X = i/N$, with $i = 1, 2,...,N$ for respondents who have the lowest, the second lowest,…, and the highest recorded values of X, respectively (O'Donnel et al., 2008).

decomposed using Eq. (2), but now with CI_k representing the CI of variable k with respect to lifestyle or health variable y.

Depending on the nature of the two variables from which to calculate it, the standard CI_s (Gini$_s$) in Eq. (1) may possess a few undesirable properties. As these limitations – which are being discussed in Erreygers (2009a, 2009b) – are relevant in this study, we therefore choose to instead use the following, recently developed version of the CI (Gini);

$$CI = 4 \frac{\mu}{b_y - a_y} CI_s, \tag{3}$$

where b_y and a_y are the upper and lower limits of y, and CI_s (Gini$_s$) is the standard version of the CI (Gini) in Eq. (1) (Erreygers (2009a, 2009b)).[7] The decomposition formula for this version of CI (Gini) is obtained by scaling Eq. (2) similarly, i.e. by $4\mu/(b_y - a_y)$. While the actual index estimates are sensitive to whether one uses the CI (Gini) versions in Eq. (1) or Eq. (3), the corresponding decomposition analysis is invariant to which version is being used, i.e. the percentage contribution of different variables to indices of inequality are identical in the two versions.

3. Correlates of lifestyles and health – multivariate probit model results

We start comparing patterns of inequality in lifestyles and health by looking at results of the multivariate probit model, which are reported in Table 2.[8] The key results are as follows.

First, controlling for a large set of potentially confounding factors, clear education gradients exist in all four lifestyle variables, while income is significantly associated only with PA and in part FV. For our health variables, Non-obese and SAH, the opposite seems to be true; clear income gradients are present, but the effects of education are less clear, with only the association between College degree and SAH being statistically significant. In general, and not surprisingly, the association between education and health, and between income and lifestyles, becomes stronger as we remove control variables from the probit models in Table 2.[9] What remains, however, is that, in relative terms, education seems to be more important than income in predicting healthy lifestyles, while the opposite seems to be true for health.

Second, while Table 2 shows that there are several significant effects of occupational status on different lifestyle and health variables, one association clearly stands out; individuals on social security are 41.6 percentage points less likely than others to report being in good health. As noted in Footnote 4, in this study we are generally not able to establish causal effects, as our data are cross sectional. This limitation seems to be particularly relevant in this example, as we suspect that the strong negative correlation that exists between social security status and SAH is mainly due to the effect of poor SAH on social security status, and not *vice versa* (as modeled in Table 2). What is probably more interesting in this context is the extent to which social security status is responsible for the strong relationship that exists

---

[7] We follow the delta method described in O'Donell et al. (2008:103) with some adjustments to obtain CIs (Ginis) according to Eq. (3). Also, since our socioeconomic indicators income and education are categorical, we follow the example of e.g. Chen and Roy (2009) by giving equal fractional rank r to ties (their average fractional rank), rather than sort people with equal incomes or education randomly, or by other variables than income or education. Further technical details on calculating and decomposing Ginis and CIs are provided in Footnote 10.

[8] Full results of the multivariate probit model in Table 2, including original parameter estimates and their robust standard errors, are available upon request. The marginal probabilities in Table 2 are average partial effects, i.e., marginal probabilities of each regressor have been calculated for each individual in the sample. The sample means of these individual level calculations represent the average partial effects. Standard errors of these average partial effects are obtained by combining the original probit parameters and their robust standard error, using the delta method.

[9] Results not shown here due to space considerations.

between income and SAH, that is, income and SAH might be strongly associated partly because poor SAH make individuals exit the labor force prematurely, which in turn affect their incomes negatively due to a shift from earning wages to being on social security. Evidence on such mechanisms of 'reverse causality' from health to income, in particular in late midlife, is provided in for example Case and Deaton (2005) and van Kippersluis et al. (2010).

Table 2. Multivariate probit model for lifestyles and health – marginal probabilities

	(1) PA	(2) FV	(3) FISH	(4) NSMOKE	(5) NOBESE	(6) SAH
Age 35-44	0.002	*0.052*	**0.067**	***-0.051***	***-0.041***	-0.074
Age 45-54	**0.002**	**0.078**	**0.119**	**-0.086**	-0.015	**-0.082**
Age 55-64	*0.045*	**0.112**	**0.186**	-0.033	0.002	**-0.135**
Age 65-74	**0.088**	**0.149**	**0.205**	0.034	0.016	**-0.111**
Female	**0.068**	**0.183**	**0.044**	**-0.029**	0.014	0.011
Household has kids	*-0.031*	-0.015	**0.044**	0.009	0.007	***0.033***
Widow	0.037	*-0.053*	-0.052	-0.017	0.033	*0.043*
Divorced	***0.049***	-0.028	**-0.087**	**-0.098**	0.022	0.010
Single	***0.055***	**-0.060**	**-0.082**	***-0.042***	-0.026	0.001
High school	0.029	0.014	0.001	-0.005	-0.020	0.000
Some college	**0.089**	***0.056***	**0.050**	**0.072**	-0.022	0.030
College degree	**0.113**	**0.096**	**0.079**	**0.127**	0.012	***0.045***
Income quartile 2	**0.048**	0.011	-0.002	0.011	0.015	**0.061**
Income quartile 3	**0.065**	0.014	0.026	*0.029*	**0.037**	**0.113**
Income quartile 4	**0.099**	***0.046***	0.011	0.025	**0.040**	**0.123**
Skilled manual	*0.031*	-0.020	0.002	-0.013	0.017	-0.002
Unskilled manual	-0.015	***-0.063***	-0.037	**-0.099**	-0.004	**-0.074**
Social security	0.015	*-0.040*	0.006	**-0.090**	**-0.050**	**-0.416**
Other occupation	**0.055**	-0.009	0.014	0.011	0.004	**-0.050**
Pay in installments	-0.018	***-0.034***	**-0.042**	**-0.039**	**-0.073**	-0.010
Life insurance	0.019	-0.003	0.004	0.018	-0.002	***0.023***
Self control	***0.039***	*0.031*	0.007	**0.065**	0.000	**0.060**
Childhood ec. aver.	-0.010	***0.034***	0.018	**0.044**	0.008	0.019
Childhood ec. rich	0.000	0.017	0.023	*0.029*	**0.034**	*0.039*
Mother high school	0.001	0.022	-0.006	-0.015	***0.026***	0.015
Mother college	-0.013	***0.058***	0.021	***0.042***	0.058	*0.053*
Father high school	-0.009	0.019	0.014	0.004	0.001	-0.017
Father college	0.031	***0.041***	0.004	0.009	0.017	0.003
2007	0.002	***0.031***	**0.051**	0.017	**-0.032**	-0.004
2009	0.012	**0.089**	*0.020*	**0.033**	**-0.040**	-0.013
Residual correlations						
PA		0.216	0.085	0.253	0.136	0.196
FV			0.181	0.213	-0.025	*0.050*
FISH				0.073	0.044	*0.058*
NSMOKE					-0.113	0.150
NOBESE						0.296
N						7738
Log-likelihood						-24158.96

Notes: This multivariate probit model was estimated using the Stata module mvprobit (with 90 draws). Sample weights were applied. Marginal probabilities in **bold***,* ***bold italics*** *and italics are statistically significant at the 99 %, 95 % and 90 % levels, respectively. The marginal probabilities represent average partial effects (see Footnote 8 for details). See Table 1 for definitions of relevant reference categories.*

Third, our proxies for time preferences and self control are in several cases significantly related to lifestyles and health. However, no clear systematic patterns stand out with respect to these variables. For example, the results in Table 2 do not indicate that psychological traits are more important in predicting lifestyle choices than health outcomes, or *vice versa*.

Fourth, a similar result holds for the role of childhood circumstances, that is, we are not able to identify any systematic differences between lifestyle and health variables with respect to for example the impact of parental education. Among the included variables for childhood circumstances, being raised by a mother who had completed some form of university level education seems to be most important, as this variable is statistically significant in two out of four lifestyle equations and in both health equations, with the range of marginal probabilities being 0.042–0.058. In fact, for the two health variables, and in particular non-obesity, the education of the respondent's mother is more important than his or her own education in predicting health.

Finally, the residual error terms of the different equations in Table 2 are in several cases strongly correlated; 11 out of 15 cross-equation correlation coefficients are statistically significant at the 99 % level (of which 11 at the 99 % level), and 7 out of these are correlated in the excess of 0.150. Thus, in general, controlling for all the regressors in Table 2, there tend to exist other, unobserved characteristics which make individuals systematically choose healthy lifestyles and have good health, or *vice versa*. The most notable exception to this pattern in Table 2 is found between non-smoking and non-obesity, where the correlation coefficient is negative, at -0.113. However, this particular result is not surprising, as there is evidence to suggest that smoking is associated with lower body weight through affecting ones appetite and metabolic rate (Chiolero et al., 2008). The correlation matrix in Table 2 suggests that the strongest correlation of unobserved individual characteristics exists between non-obese and SAH (0.296), i.e., between our two health variables, while physical activity is the lifestyle variable which is most closely associated with these two health measures, again in terms of unobserved characteristics. While the correlation structure for unobserved characteristics between the different lifestyle variables and non-obesity is mixed, healthy lifestyles and good SAH are always positively related, and significantly so.

4. Decomposing total inequality and socioeconomic inequality in lifestyles and health

We turn next to inequality in lifestyles and health as measured by Gini and Concentration indices. The results of interest are presented in Table 3. The row "Gini (G)/CI_{educ}/CI_{inc}" report the actual index estimates, calculated according to Eq. (3), for respectively total inequality, education-related inequality and income-related inequality in lifestyles and health. The remaining rows in Table 3 report results from the corresponding decomposition analyses (Eq. 2), that is, these rows indicate the percentage contribution of each regressor, or group of regressors (in bold), to the Gini, CI_{educ} and CI_{inc} for each lifestyle and health variable.[10]

---

[10] We follow the procedure of Balia and Jones (2008); since all our outcome variables are binary, we base our Gini calculations on predicted probabilities rather than observed outcomes, with individual predictions being based on results of the multivariate probit model in Table 2. This procedure ensures that we get sufficient variability in the outcome variables for which to calculate the Ginis, but this comes at a cost; predicted probabilities are additive in the regressors, and thus only the deterministic part of the decomposition equation (Eq. 2) can be calculated (Balia and Jones, 2008). Thus, although the percentage contributions per column in Table 3 sum to one hundred (summing groups of regressors, in bold), this only reflects the explained part of total inequality. Similar to econometric models for individual behavior and health, the unexplained residuals of variation, or here, the unexplained residuals of inequality, are typically large, and this must be taken into account when reviewing these results. For consistency, we also calculated the education- and income-related CIs in Table 3 using predicted probabilities, although in principle, these could be calculated using observed outcomes.

Table 3. Decomposing Gini indices of total inequality and CIs of socioeconomic inequality in lifestyles and health – percentage contributions

	PA			FV			FISH			NSMOKE			NOBESE			SAH		
	G	CI_{ed}	CI_{inc}	G	CI_{ed}	CI_{inc}	G	CI_{ed}	CI_{inc}	G	CI_{ed}	CI_{inc}	G	CI_{ed}	CI_{inc}	G	CI_{ed}	CI_{inc}
Demographics	**37.0**	**-15.9**	**-15.0**	**58.9**	**-15.1**	**-9.8**	**76.1**	**-56.1**	**5.0**	**18.5**	**-1.4**	**-0.4**	**13.2**	**-10.8**	**-12.6**	**18.1**	**10.7**	**-6.9**
Age	12.7	-12.0	-0.8	8.2	-16.7	-2.0	64.8	-68.6	-6.6	9.8	-2.5	-14.0	8.1	-8.3	-4.1	15.5	9.7	-2.5
Female	14.0	0.6	-7.0	47.8	1.5	-20.4	4.3	0.7	-8.7	1.7	-0.2	2.9	1.7	0.2	-2.7	-0.1	0.1	-0.6
Household has kids	6.1	-3.2	3.5	0.6	-1.4	1.8	-2.9	8.5	-9.6	0.2	0.6	-0.9	-0.1	1.5	-1.5	3.8	2.3	-1.9
Marital status	4.2	-1.3	-10.6	2.4	1.5	10.8	9.9	3.3	29.9	6.8	0.7	11.7	3.6	-4.2	-4.4	-1.2	-1.3	-1.8
Education	**27.1**	**82.9**	**33.2**	**12.7**	**65.3**	**31.3**	**10.3**	**124.3**	**51.5**	**41.2**	**71.2**	**43.2**	**11.6**	**37.5**	**12.9**	**7.7**	**26.1**	**8.1**
Income	**18.1**	**22.1**	**79.4**	**3.9**	**8.4**	**38.1**	**1.6**	**7.6**	**29.3**	**3.7**	**4.1**	**22.6**	**11.6**	**20.5**	**69.5**	**21.1**	**20.8**	**58.6**
Occupation	**7.9**	**-2.3**	**-10.2**	**3.6**	**10.1**	**11.9**	**3.8**	**5.2**	**-1.9**	**14.5**	**10.5**	**14.4**	**8.1**	**7.2**	**10.2**	**37.7**	**26.3**	**31.5**
Skilled manual	-0.5	-0.8	-0.4	0.6	0.5	0.3	-0.2	-0.1	-0.1	0.3	0.2	0.2	0.9	-0.9	-0.5	-0.1	0.0	0.0
Unskilled manual	1.2	1.6	0.9	2.9	6.4	4.1	2.3	7.1	4.0	6.7	6.2	5.1	0.2	0.8	0.4	2.2	5.2	2.2
Social security	-0.4	-1.3	-2.0	0.6	3.1	5.9	0.1	-1.0	-1.6	6.6	4.4	10.8	6.7	7.6	11.5	31.9	20.0	25.2
Other occupation	7.6	-1.9	-8.6	-0.4	0.3	1.6	1.6	-0.9	-4.2	0.9	-0.2	-1.7	0.3	-0.3	-1.2	3.7	1.1	4.1
Psychological traits	**5.3**	**7.1**	**7.5**	**3.7**	**4.7**	**5.5**	**3.7**	**5.4**	**6.5**	**10.2**	**6.7**	**10.6**	**17.7**	**5.3**	**5.9**	**6.8**	**6.4**	**5.2**
Pay in installments	1.3	0.8	0.9	2.1	1.3	1.9	3.5	3.2	4.1	2.4	1.0	1.9	17.8	5.6	6.4	0.2	0.3	0.3
Life insurance	0.3	1.3	2.0	0.0	-0.2	-0.3	-0.1	0.5	0.8	1.0	0.8	1.8	-0.1	-0.3	-0.5	2.1	1.1	1.3
Self control	3.7	5.0	4.5	1.6	3.6	3.9	0.2	1.7	1.6	6.7	4.9	6.9	0.0	0.0	0.0	4.5	5.0	3.6
Childhood conditions	**3.9**	**5.4**	**3.9**	**7.8**	**21.8**	**14.0**	**0.3**	**13.3**	**7.3**	**9.7**	**7.6**	**6.6**	**32.1**	**44.0**	**20.4**	**8.7**	**10.2**	**4.1**
Childhood ec. situation	0.3	0.1	-0.1	1.2	1.0	1.3	0.4	3.5	2.7	2.7	1.3	1.8	5.7	6.5	3.8	2.1	2.4	1.2
Mother's education	-0.6	-2.4	-1.1	3.4	11.1	5.7	0.4	7.0	3.4	5.7	4.8	3.4	20.9	28.7	11.4	6.5	7.7	2.6
Father's education	4.2	7.7	5.1	3.2	9.7	7.0	-0.4	2.8	1.2	1.2	1.5	1.4	5.5	8.8	5.2	0.2	0.1	0.3
Survey years	**0.8**	**0.8**	**1.2**	**9.3**	**4.7**	**9.0**	**4.1**	**0.4**	**2.3**	**2.3**	**1.2**	**3.0**	**5.6**	**-3.8**	**-6.2**	**-0.1**	**-0.6**	**-0.7**
Gini (G) / CI_{ed} / CI_{inc}	.380	.203	.179	.371	.156	.114	.377	.098	.082	.356	.250	.144	.309	.130	.123	.368	.188	.210

Notes: The Ginis and the education- and income-related CIs have been calculated and decomposed according to Eq. (3) and Eq. (2), respectively. The numbers in this table indicate the percentage contribution of each regressor (or each group of regressors) to the explained part of the decomposition formula in Eq (2). Further technical details are provided in Footnote 10.

The results in Table 3 suggest that decomposition-of-inequality analyses are very sensitive to which 'type of inequality' is being studied. While perhaps not surprising, education itself is generally a very dominant contributor to education-related CIs in lifestyles and health (mean contribution: 67.9 %), while income is similarly a dominant contributor to income-related CIs (mean contribution: 49.6 %). In contrast, if we instead focus on sources of total inequality, education and income become less important, as they explain on average 18.4 % and 10.0 % of the Ginis in lifestyles and health.

Another example of the sensitivity issue is the role of gender in fruits and vegetables, which explains as much as 47.8 % of total inequality, but only 1.5 % of CI_{educ}, which is mainly due the CI of education with respect to gender being close to zero (the second component of Eq. 2). Thus, at least in our sample, a study focusing on socioeconomic inequality in fruits and vegetables eating would probably miss out that the key target group for eating more fruits and vegetables is actually males, and not low income or education groups (although these groups are also important). As we believe that factors other than socioeconomic status are also important elements of inequality in lifestyles and health (Fleurbaey and Schokkaert 2009), we will in the following focus mainly on sources of total inequality in these variables, i.e., on decompositions of the Gini indices in Table 3.

The key contributors to total inequality are not the same across our different lifestyle and health variables. As indicated, the key contributor in fruits and vegetables is gender (47.8 %), while in fish eating, age is clearly most important (64.8 %). Education is the key factor in explaining population differences in physical activity (27.1 %) and in particular non-smoking (41.2 %), while social security status explains as much as 31.9 % of the Gini in SAH, which reflects the strong association that was found between these variables in the multivariate probit model in Table 2. Thus, social security status is the most important factor in explaining inequality in SAH, despite its relatively low mean (0.082), which, *ceteris paribus*, should reduce its impact on total inequality (see Eq. 2). Interestingly, childhood conditions seem to be particularly important in explaining population differences in body mass, with maternal education being the single most important contributor to the Gini in non-obesity (20.9 %).

The finding in the multivariate probit model in Table 2 of education being relatively more important than income in predicting lifestyles, and of income being relatively more important than education in predicting health, is partly reflected in the contribution of these two indicators of socioeconomic status to total inequality; on average, education and income explain respectively 22.8 % and 6.8 % of the Ginis in lifestyles, while the corresponding figures for SAH are 7.7 % and 21.1 %. Education and income make identical contributions to the Gini in non-obesity (11.6 % each), which is different from results of the multivariate probit model in Table 2, where income was found to be more important than education. Of course, the finding of education having a greater impact on total inequality in lifestyles than SAH does not imply that there are greater educational differences in lifestyles than SAH. What it does imply is that, (i) while both lifestyles and SAH are strongly correlated with education, and (ii) education is strongly correlated with many of the other control variables in Table 3, (iii) these other control variables are more directly associated with SAH than with the different lifestyles, which means that the direct contribution of education itself to indices of total inequality and education-related inequality is more 'attenuated' in SAH than in lifestyles.

5. Conclusions

The main purpose of this paper has been to compare sources and patterns of inequality in important health affecting lifestyle choices, on the one hand, with those in final health, on the other. The motivation for this assessment, which uses Norwegian data, has been that health

inequalities should preferably be prevented rather than treated, and thus, policies must mainly target production factors of health, including lifestyles, and not final health itself. However, the existing literature has mainly focused on sources and patterns of inequality in final health. For policy purposes, knowledge about these patterns in final health is mainly relevant to the extent that they are representative of patterns of inequality in important, underlying production factors of health.

As is standard, we find that there are clear income and education gradients in most of our lifestyle and health variables, which are (i) physical activity, non-smoking, and eating fish and fruits and vegetables, and (ii) obesity status and self assessed health. However, in a multivariate probit model that controls for basic demographics, occupational status, psychological traits, and childhood circumstances, there is considerable variation across the different lifestyle and health variables with respect to the steepness and the statistical significance of these socioeconomic gradients.

Using decomposition techniques for the Gini index of total inequality in lifestyles and health, we find that there is considerable variation across these variables with respect to sources and patterns of inequality. While education is generally an important source of inequality, and in some variables – primarily the health variables – also income, there are several cases in which other factors are much more important in explaining inequality, such as gender in fruits and vegetables eating, age in fish eating, and maternal education in obesity.

Our main conclusion is therefore that patterns of inequality in different lifestyle and health variables are heterogeneous, and thus, patterns of inequality in health variables are not necessarily representative and relevant for patterns of inequality in their underlying production factors, including lifestyles. While one could argue that population differences in lifestyles and health by socioeconomic status are particularly problematic and unjust, and thus rightfully achieves almost all attention in the literature on health inequalities, we agree with Fleurbaey and Schokkaert (2009) in that more attention should also be given to other, perhaps more important sources of inequality, such as age and gender in the examples above, rather than simply be labeling these as 'acceptable' or 'unavoidable' sources of inequality. For example, it is clearly possible to avoid having gender differences in fruits and vegetable eating, and in fact, achieving it is a highly workable policy goal.

References
Balia, S., Jones, A.M., 2008. Mortality, lifestyle and socio-economic status. Journal of Health Economics 27, 1–26.
Cappellari, L., Jenkins, S.P., 2003. Multivariate probit regression using simulated maximum likelihood. The Stata Journal 3 (3), 278–294.
Case, A., Deaton, A., 2005. Broken down by work and sex: how our health declines. In: Wise, D. A. (Ed.), Analyses in the economics of aging. Chicago: Chicago University Press, p. 185–205.
Case, A., Fertig, A., Paxson, C., 2005. The lasting impact of childhood health and circumstance. Journal of Health Economics 24(2), 365–389.
Chen, Z., Roy, K., 2009. Calculating concentration index with repetitive values of indicators of economic welfare. Journal of Health Economics 28, 169–175.
Chiolero, A., Faeh, D., Paccaud, F. Cornuz, J., 2008. Consequences of smoking for body weight, body fat distribution, and insulin resistance. Am J Clin Nutr 87, 801–809.
Connor Gorber, S., Tremblay, M., Moher, D., Gorber, B., 2007. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obesity Reviews 8(4), 307–326.
Contoyannis, P., Jones, A.M., 2004. Socio-economic status, health and lifestyle. Journal of Health Economics 23, 965–995.

Costa-Font, J., Gil, J., 2008. What lies behind socio-economic inequalities in Spain? A decomposition approach. Food Policy 33, 61–73.

Cutler, D.M., Lleras-Muney, A., 2010. Understanding differences in health behaviors by education. Journal of Health Economics 29, 1–28.

Erreygers, G., 2009a. Correcting the Concentration Index. Journal of Health Economics 28, 504–515.

Erreygers, G., 2009b. Can a Single Indicator Measure Both Attainment and Shortfall Inequality? Journal of Health Economics 28 (4), 893–901.

Fleurbaey, M., Schokkaert, E., 2009. Unfair inequalities in health and health care. Journal of Health Economics 28(1), 73–90.

Heckman, J.J., 2007. The economics, technology, and neuroscience of human capability formation. Proceedings of the National Academy of Sciences 104, 13250–13255.

Marmot, M., Friel, S., Bell, R., Houweling, T.A., Taylor, S., 2008. Closing the gap in a generation: health equity through action on the social determinants of health. Lancet 372, 1661–1669.

O'Donnell, O., van Doorslaer, E.,Wagstaff, E., Lindelow, M., 2008. Analyzing health equity using household survey data. Washington, D.C: World Bank Institute.

OECD, 2008. Growing unequal ? Income distribution and poverty in OECD countries, Paris. (www.oecd.org/els/social/inequality / www.oecd.org/els/social/inegalite).

Rosa Dias, P., 2009. Inequality of opportunity in health: evidence from a UK cohort study. Health Economics 18(9), 1057–1074.

Rosa Dias, P., 2010. Modeling opportunity in health under partial observability of circumstances. Health Economics 19, 252–264.

Trannoy, A., Tubeuf, S., Jusot, .F, Devaux, M., 2010. Inequality of opportunities in health in France: a first pass. Health Economics 19(8), 921–938.

Vallejo-Torres, L., Morris, S., 2010. The contribution of smoking and obesity to income-related inequalities in health in England. Social Science & Medicine 71, 1189-1198.

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 Kippersluis, H., O'Donnell, O., van Doorslaer, E., van Ourti, T., 2010. Socioeconomic differences in health over the life cycle in an egalitarian country. Social Science & Medicine 70, 428–438.

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 (1), 207–223.

World Health Organization, 2000. Obesity: preventing and managing the global epidemic. WHO technical report series no. 894, Geneva, Switzerland.

World Health Organization, 2003. Diet, nutrition and the prevention of chronic diseases. WHO technical report series no. 916, Geneva, Switzerland.