Determinants of Inequality in BMI

Leonie Sundmacher* and Stephen Morris

Health Economic Research Group, Brunel University

Abstract

Aim: Previous studies have analysed the determinants of body mass index (BMI) in England. The aim of this study is to analyse the determinants of explained inequality in BMI, with a focus on socioeconomic factors.

Methods: We use a regression-based approach which has been applied previously to analysing inequality in income and decompose the explained sum of squares of variation in BMI into different sources. The method is based on Shorrocks (1982, 1983) and Fields (2004). The sample is stratified by gender.

Data: The data for the analysis was taken from the 2005 round of the Health Survey for England (HSE). The sample consisted of non-pregnant women and men age 16 years and older (n=10222).

Results: Age and area of residence were the strongest contributors to explained inequality in the regression analysis. Socioeconomic factors taken together (income, education, employment status, education, social class, car ownership and area deprivation) contribute 13% to explained inequality in BMI in male compared to approximately 31% in female.

Conclusion: In summary, our paper suggests that socioeconomic status plays an important role in the explanation of inequality in BMI in females, and to a lesser extent in males.

1. Introduction

England faces a severe obesity crisis. In 2005, the prevalence rates of overweight and obesity in England were 42.6% and 22.1%, respectively, for men and 32.1% and 24.2%, respectively, for women (OECD Health Data 2007). This nearly threefold increase in obesity since 1980 (Lister, 2005) imposes major physical and economic burden on the English population; in 2001, it has been estimated that over 34,000 deaths in England, or about seven percent of all deaths, could be attributed to obesity (House of Commons Health Committee, 2004). In the same year, obesity accounted for over 15 million days of sickness, costing the English economy £1,322 million in lost earnings (House of Commons Health Committee, op cit).

* *Correspondence to: Leonie.Sundmacher@brunel.ac.uk*

The recently released Foresight report on obesity states that the scope of the crisis requires a new future role of government in regulation, investment and promotion of good health. In this context, the report stresses the important role of socioeconomic factors in the obesity crisis and warns of a continued polarisation into the "junk food eating, less-educated poor" on the one hand and functional food eating, better informed higher classes" on the other (Foresight Report, 2007). The report therefore suggests that improving education for the poorest group may help reduce obesity. However, empirical evidence on what determines the unequal distribution of body mass in the UK population is rare and experience on interventions to reverse obesity are largely absent. Research is required to fill this gap and improve the understanding of social determinants of obesity and investigate their relative importance in the crisis. This information may then help decision makers to evaluate the likely impact of programmes that aim to improve socioeconomic conditions in order to curb the epidemic.

In the light of the above, this paper aims to analyse the determinants of inequality in body mass index (BMI) in England and Wales. BMI is measured as weight in kilograms divided by height in metres squared. The BMI range generally considered to be healthy is 20 to 25kg/m^2. Those with a BMI below 20kg/m^2 are underweight, and those with a BMI in the range 25 to 30kg/m^2 are overweight. Obesity is usually defined by a BMI over 30kg/m^2. We first identify the extent of BMI inequality using a variety of standard inequality measures applied to individual level health survey data. We then identify the determinants of BMI using a regression-based approach which has been applied previously to analysing inequality in income. The approach is based on decomposing the explained sum of squares of variation in BMI into its different sources. The method is based on Shorrocks (1982, 1983) and Fields (2004). Using this approach, our analysis focuses on the contribution of socioeconomic factors to the explained component of inequality in BMI. Our analysis is conducted using individual level data from the 2005 Health Survey for England.

The next section reviews the decomposition methods and existing literature on inequality in BMI, focusing on socioeconomic factors. Section 3 presents Fields' method used to decompose variation in BMI into different sources. Sections 4 and 5 discuss the data used to perform the analysis and the estimation process respectively, section 6 presents the results and section 7 concludes.

2. Decomposing inequality

Research in decomposition of inequality is rooted and driven by research applied to income inequality. In early studies, income decomposition by population groups has been the leading approach to quantify how factors affect inequality. Following this approach, a sample is divided into categories (e.g., high and low education attainment) and the level of inequality is calculated within each subsample and between the means of each of the sub samples using a number of standard measures including generalised entropy measures and Atkinson indices. The pitfalls with this approach are that the decomposition can only be carried out over discrete categories, even if continuous explanatory factors are plausible and that inequality with respect to multiple factors is difficult to consider since the number of sub groups increases multiplicatively with the number of categories for each explanatory factor (Morduch and Sicular, 2002). To overcome these pitfalls Shorrock (1982) developed a method that decomposes inequality by sources of 'factor components', rather than based on population groups. Specifically in the context of income inequality, he proposed to write inequality measures *I* as the weighted sum of incomes y from k different sources:

$$I(y) = \sum_i a_{ik}(y) y_{ik} \qquad [1]$$

where $a_{ik}(y)$ denote the weights for income source k. Dividing [1] over I(y), then gives the proportional contribution of income source *k* to overall inequality, denoted by $s_k = \frac{\sum_i a(y) y_{ik}}{I(y)}$.

Shorrocks (1982) then pointed out that this decomposition method yields an infinite number of potential decomposition rules. This is because the weights $a_{ik}(y)$ can be chosen in many ways, thus altering the proportional contribution of each factor. He therefore proposed additional restrictions on the choice of weights in order to reduce the number of potential decomposition rules (listed in appendix I). The two most important conditions are (i) that if all individuals have the same value in y for the k'th factor, then the contribution of this factor to inequality is zero, and (ii) if the distribution of y for factor k=1 is only a permutation of the distribution of y for factor k=2, and if k=1,2 are the only two components in the decomposition, then they receive the same share in the decomposition. Imposing these restrictions, he arrives at the unique decomposition rule:

$$s_k = \frac{\text{cov}(y_k, y)}{\text{var}(y)} \qquad [2]$$

Murdoch and Sicular (2002) and Fields (2004) extend Shorrocks approach to a regression based decomposition of inequality. They express household income as linear function of explanatory variables and use the regression coefficients to calculate the decomposition components for all variables included in the model. The regression-based decomposition has the advantages that (1) it yields an exact allocation of contributions to the identified factors, (2) it provides measures of uncertainty around the values of *s* which are part of standard regression analysis and (3) it allows for the analysis of multiple factors. In addition - choosing a different path of reasoning - Fields arrives at the same equation like Shorrocks (1982) in [2] and therefore considers the restrictions of the choice of weights.

In health economic research, van Doerslaer (2002) showed how the concentration index which is typically used to measure income-related inequalities in health or health service use can be decomposed. Doerslaer et al used as well regression-based decomposition methods to explore how correlates of health contribute to total income-related inequality in health in two ways: first, via their impact on health measured as health elasticity and second, via their share of unequal distribution across income, measured in terms of the concentration index.

This paper uses Fields' (2004) approach to decompose inequality in BMI. However, it should be noted that Fields' approach only differs from the decomposition of the concentration index in such that it directly decomposes inequality in BMI (not income-related inequality in BMI) measured as explained variance within a regression framework. Thus, it is a decomposition method intended to supplement regression analysis. The approach is explained in section 3.

Evidence from econometrics studies

To our knowledge there exist only two studies which explore the relationship between socioeconomic factors and inequality in weight-height relation. Both studies employ the concentration index, and one uses decomposition methods. Zhang and Wang (2004) assess socioeconomic inequality in the distribution of obesity in the US. Comparing socioeconomic inequality across gender, age and ethnicity, they find a stronger inverse association between socioeconomic status and obesity in female compared to men and a higher prevalence of obesity among whites compared to other ethnic groups. Costa-Fonta and Gil (2005) explain socioeconomic inequalities in obesity in Spain. They decomposed income-related obesity into sources and find that education explains most of the variation in obesity, followed by income, physical exercise and region of residence.

3. Decomposing inequality in BMI using Fields' method

We begin with the estimation of the factors associated with BMI. BMI is regressed on a range of covariates using a standard least squares regression model of the form:

$$Y_i = \beta_0 + \sum_{k=1}^{K} X_k \beta_k + \varepsilon \qquad [3]$$

where Y_i is the BMI of individual i, X_k is a vector of variables $X_1, X_2,...X_K$ thought to determine BMI (there are k = 1, 2, … K variables included in X_k) and β_k is a vector of coefficients $\beta_1, \beta_2...\beta_k$ pertaining to each variable $k.\varepsilon$ is an error term with a mean value of zero and a variance of unity and β_0 is the intercept term. The estimated coefficients are denoted by

$$(\hat{\beta}_0, \hat{\beta}_{1....}\hat{\beta}_K) \quad [4]$$

and the residual term is given by

$$\hat{\varepsilon}_{ik} = Y_i - \hat{\beta}_0 - \sum_{k=1}^{K} X_{ik}\hat{\beta}_k, i = 1,..., n \quad [5]$$

We then decompose explained variance in BMI using Fields (2004) method. We therefore let $s(X_k)$ denote the share of the variance in BMI attributable to the *k*th determinant holding all other determinants constant and the model R^2 is the proportion of variance explained by all determinants X_k taken together. To deduct the decomposition, Fields (2004) first takes the variance of the left and right hand sides of equation [3], which is written as

$$\sigma_y^2 = \sum_{k=1}^{K} \mathrm{cov}\left[X_k\hat{\beta}_k, Y\right] + \mathrm{cov}\left[\hat{\varepsilon}, Y\right] \quad [6]$$

Dividing [6] by the variance of *Y* then yields

$$1 = \frac{\sum_{k=1}^{K} \mathrm{cov}\left[X_k\hat{\beta}_k, Y\right] + \mathrm{cov}\left[\hat{\varepsilon}, Y\right]}{\sigma_y^2} = \sum_{K=1}^{K} s_k(X_k) + s(\hat{\varepsilon}) \quad [7]$$

The equation partitions the full variance of *Y* into the share that is explained by the covariance between each of the *X* factors and the *Y* values. Fields calls the proportions denoted by $s(X_k)$ ”relative factor inequality weights”. Note that $\frac{\sum_{k=1}^{K} \mathrm{cov}\left[X_k\hat{\beta}_k, Y\right]}{\sigma_y^2} = \sum_{K=1}^{K} s_k(X_k)$ is the model R^2 .

Dividing the individuals' *s* weights for each *k* by the model R^2 gives the share of each factor in the explained variation of the linear regression. Formally, this is given by

$$p_k(X_k) \equiv \frac{s_k(X_k)}{R^2{}_y} \equiv \frac{\mathrm{cov}\left[X_k\hat{\beta}_k, Y\right]}{\sum_{k=1}^{K}\mathrm{cov}\left[X_k\hat{\beta}_k, Y\right]} \qquad [8]$$

Furthermore, Fields shows that under six decomposition conditions (listed in appendix I), the s-weights and p-weights are the same for any measure of dispersion that is continuous, symmetric, and take value zero when all Y are identical. (namely the Gini Coefficient, Theil index, and Atkinson index).

Interpretation of the decomposition results

Two points should be bear in mind when interpreting the decomposition results: first, Fields approach decomposes the predicted value of *Y* rather than the actual value of *Y*. Thus, using Fields' approach, we quantify the relative importance of determinants of explained inequality in Y. Second, Fields (2004) himself pointed out that the p weights can take negative values. He explained this result as follows: the covariance between $X_k\hat{\beta}_k$ and Y can be expressed as $\mathrm{cov}[X_k\hat{\beta}_k, Y] = \hat{\beta}_k \,\mathrm{cov}[X_k, Y]$, and when we regress of Y on X_k we obtain the simple regression coefficient $\hat{\beta}_{X_kY} = \frac{\mathrm{cov}[X_k, Y]}{\sigma^2}$. It follows that the p-weights can be written as $p(X_k) = \frac{\hat{\beta}_k \hat{\beta}_{X^k,Y}}{R^2}$. This implies that a negative value arises whenever the two beta coefficients have opposite sign i.e. whenever controlling for multiple factors within a regression framework would reverse the sign from the simple regression.

4. Data and variables

In this section we describe the data and variables used to investigate BMI inequality using Fields' method. The data came from the 2005 round of the Health Survey for England (HSE). The HSE is a nationally representative survey of individuals aged 2 years and over living in England. A new sample is drawn each year and respondents are interviewed on a range of core topics including demographic and socioeconomic indicators, general health and psychosocial indicators, and use of health services. Additionally, there is a follow up visit by a nurse at which various physiological measurements are taken, including height and weight.

BMI is computed for each respondent from height and weight values obtained during the nurse visit and reported in the HSE. One useful feature of the HSE is that height and weight are collected by a nurse and are therefore not prone to reporting bias.

The total sample size in the 2005 HSE is 13,297. We only include non-pregnant individuals over 15 years of age in our analysis (children and pregnant women were excluded because their measurement of obesity does not relate to the standard BMI categories). Excluding by age and pregnancy reduces the sample to 10,222. Of these individuals, 8,384 have a valid BMI measure from the nurse visit. Previous research suggests that the determinants of BMI and obesity are different for males and females (Sundquist and Johansson 1998; Zagorsky 2005; Zhang and Wang 2006; Fernaldi 2007); we therefore stratify the analysis by gender; 4,512 individuals in our final sample are female and 3,872 are male.

We include the following covariates in our BMI regressions: income (measured as equivalised household income) and income squared; a cubic function of age; education attainment; ethnicity; marital status; economic activity; housing tenure; social class of household response person ; number of infants aged 0–1 years living in the household; number of children aged 2–16 years living in the household; total number of people (infants plus children plus adults) living in the household ;area deprivation measured using the Index of Multiple Deprivation 2004 (ODPM, 2004); number of portions of fruit consumed on average each day, and number of portions of fruit consumed squared; number of cars owned in the household; month of interview; and, regional (strategic health authority) indicators. In the analysis, we take the view that income, education, employment status, education, social class, car ownership and area deprivation define socioeconomic status whereas marital status, tenure, number of children, persons and infants in the household and the area of residence are family variables. However, we acknowledge that the boundaries between socioeconomic and family variables are not clear cut.

To allow for the possibility that items are not missing at random we include dummy variables for all missing values of the covariates.

5. Results

Descriptive statistics

The sample distribution of BMI is shown in Table 1. Twenty nine percent of males and 35% of females have a BMI within the normal healthy range while 26% of women and 24% of men are classified as obese. Clearly, the majority of the population, namely 64%, is either obese or overweight.

Table 1. BMI categories by gender

BMI (kg/m²)	***Males n=3872*** ***45.19% of sample***	***Female n=4512*** ***54.81% of sample***
<20	3.28	5.36
20-25	28.54	34.8
25-30	44.47	33.91
30-35	18.34	16.34
35-40	4.55	6.41
>40	0.83	2.68
>35	5.38	9.09
>30	23.71	25.93
>25	68.19	59.34

Inequality in BMI

Inequality in BMI is quantified using three commonly used measures in Table 2. The inequality measures are the Generalized Entropy (*GE*) class of inequality indices, the Atkinson (*A*) class, and the Gini Coefficient (see Appendix I for further details about these measures). We present inequality indices for the whole sample and then for sub-groups.

Table 2. Inequality in BMI by gender

	GE(-1)	***GE(0)***	***GE(1)***	***GE(2)***	***A(0.5)***	***A(1)***	***A(2)***	***Gini***
Whole Sample	0.0164	0.0163	0.0166	0.0172	0.0082	0.0162	0.0317	0.1016
Male/ Female								
Male	0.0129	0.0127	0.0127	0.0129	0.0063	0.0126	0.0252	0.0894
Female	0.0193	0.0195	0.0200	0.0209	0.0098	0.0193	0.0372	0.1113
Between Group	0	0	0	0	0	0.00001	0.00006	
Male								
1st Income Quintile	0.0140	0.01371	0.01355	0.01359	0.00679	0.01362	0.02742	0.09253
2nd Income Quintile	0.01401	0.0138	0.0138	0.014	0.00688	0.01371	0.02725	0.09294
3rd Income Quintile	0.01352	0.01343	0.01351	0.01378	0.00671	0.01334	0.02634	0.09214
4th Income Quintile	0.01115	0.0111	0.0112	0.01144	0.00556	0.01104	0.02182	0.08302
5th Income Quintile	0.01095	0.01081	0.0108	0.0109	0.00539	0.01075	0.02142	0.08269
Between Group	0.00002	0.00002	0.00002	0.00002	0.00001	0.00003	0.00007	
Female								
1st Income Quintile	0.02256	0.02275	0.02351	0.02493	0.01149	0.0225	0.04317	0.1195
2nd Income Quintile	0.01825	0.01805	0.01816	0.0186	0.00901	0.01789	0.03521	0.10725
3rd Income Quintile	0.01863	0.01878	0.01926	0.0201	0.00947	0.01861	0.03593	0.10976
4th Income Quintile	0.01754	0.01769	0.01811	0.01882	0.00891	0.01753	0.0339	0.1068
5th Income Quintile	0.01563	0.0158	0.01622	0.01693	0.00797	0.01567	0.0303	0.10021
Between Group	0.00033	0.00033	0.00033	0.00033	0.00015	0.00029	0.00049	

In the whole sample, the Gini coefficient takes a value of approximately 0.1 indicating clear inequality between peoples body mass index. Furthermore, the Theil index GE(1) and the half square of the coefficient of variation index GE(2), which are more sensitive to inequality at the top of the distribution, have higher values indicating higher inequality at the top end of the BMI distribution. Across all measures, in-between-group inequality tends to zero for gender differences; this result is surprising but seems to be consistent with the finding that means (27.24 for females and 27.17 for males) and medians (26.23 for females 26.86 for males) are similar in the distributions for men and women. Stratifying the sample by gender and income quintiles, we find that inequality in body mass values decreases as income increases. This is true for both, men and women, while the level of inequality is generally higher among female.

Decomposition results

Following the brief description of inequality in BMI above, we now decompose inequality by its sources, with focus on socioeconomic factors. As we noted above Fields' method which is based on Shorrocks' is applicable to a number of inequality measures that meet certain criteria which are listed in appendix I.

The full results of the regression models used in the decomposition are in Appendix II; separate results are presented for men and women. We report the coefficient on each covariate and the t-value. We also report the *s* and *p* weights for every covariate. In men, BMI is significantly associated with income, age, educational attainment, marital status, housing tenure, social class of household response person, number of adults living in the household, area deprivation, number of infants living in the household, region of residence, month of interview, and car ownership. In females, the factors that are significantly associated with BMI are income, age, education, employment status, tenure, social class of household response person, the number of persons and children living in the household, area of residence and car ownership.

The results of the decomposition analysis, based on the full results in Appendix II are in Table 3, which reports the percentage contribution of each variable to the explained sum of squares, i.e., the sum of the *p* weights per determinant.

Table 3. Percentage contribution of variables to the explained sum of squares

Dependent variable BMI	***Percentage per variable Women***	***Percentage per variable Men***

Socioeconomic Factors		
Income	0.069893	0.003526
Education	0.078079	0.019642
Employment Status	0.072568	0.000373
Social Class	0.04915	0.05653
Number of cars owned	0.020117	0.037132
Area Deprivation	0.022213	0.004721
Total	**0.313585**	**0.12958**
Other Factors		
Age	0.389237	0.540308
Ethnicity	0.001565	0.007656
Marital Status	0.042805	0.118578
Tenure	-0.0089	-0.00358
Number of Children in Household	0.046343	0.009469
Number of Persons in Household	0.002997	0.070205
Number of Infants in Household	-0.00119	0.003706
Area of Residence	0.190783	0.096065
Month of the Interview	0.019488	0.021047
Daily portions of fruit	0.000256	0.009378
Missing values	0.004591	0.005248
$\mathbf{R^2}$	**0.1138**	**0.1747**
N	3388	2923

In the regression model for men the covariates explain approximately 17% of the variation in BMI. In terms of decomposition results, we find that age contributes the biggest share of explained variation, namely 54%. This is followed by marital status (11%) and then area of residence (9%). The factors we defined to explain socioeconomic status (income, education, social class, employment status, car ownership and the degree of deprivation in the area of residence) account for about 11% of the explained variation.

In women the covariates explain around 11% of the variation in BMI. Age again accounts for the largest share of the explained variation, at 39%. Marital status adds 4% and children and persons living in the household 4% and 2% respectively. The area of residence is again a good explanatory factor with a share of about 20% of the explained variation. Income, education, social class, economic status, area deprivation and car ownership account for about 31% of the variation in BMI, which is a significantly higher proportion than for men. However, as noted above, Fields approach decomposes the explained sum of square rather than overall inequality in BMI. Thus, when comparing the impact of socioeconomic factors across gender, the different R squared values of the model for men and women (17% and 11% respectively) have to be taken into account.

6. Concluding remarks

In this paper, we used Health Survey for England data to analyse the factors associated with inequality in BMI. We use a new regression-based method derived by Fields (2004) to calculate the contribution of different factors to explained variation in BMI with a focus on socioeconomic factors. For men and women, we regress a full set of covariates on BMI, and then decompose the variation in BMI into different sources. Age and area of residence were the strongest contributors to explained inequality in the regression analysis. The combined impact of socioeconomic factors captured by income, education, employment status, education, social class, car ownership and area deprivation explain 31% of the correlation coefficient R squared in female compared to only 13% in male. However, when interpreting the results it has to be borne in mind that Fields' (2004) method decomposes explained inequality in BMI. This is different from total inequality in BMI and we should interpret our results in the light of the explained proportion of variance in the regression analysis.

In summary, our paper suggests that socioeconomic status plays an important role in the explanation of inequality in BMI. In terms of gender differences, we find that socioeconomic differences contribute more to explained variation of BMI in female. An advantage of Fields' approach would be that we could analyse a direct causal relationship between factors and BMI. However, it is difficult to claim a causal relationship between determinants and BMI in our analysis as we do not control for potential endogeneity. We suggest that further research is undertaken to investigate this issue.

References

Costa-Fonta J and Gil J (2005): Are there Socio-Economic Inequalities in Obesity in Spain? FEDEA Discussion Paper.

Fernald C.H (2007): Socioeconomic Status and body mass index in low-income Mexican adults, Social Science and Medicine, 2030 – 2042

Fields G: Regression-Based Decomposition (2004): A New Tool for Managerial Decision-Making. Discussion Paper. Department for Labour Economics, Cornell University.

Fields, G.S. and Mitchell, J. (1999): Changing income inequality in Taiwan: A decomposition analysis. In T.N. Srinivasan and G. Saxonhouse (Eds.), Development, Duality, and the International Regime: Essays in Honor of Gustav Ranis. Ann Arbor: University Michigan Press

Foresight Report (2007): Tackling Obesity – Future Choice. Commissioned by the Government Office for Science. Available under http://www.foresight.gov.uk/Obesity/obesity_final/20.pdf

Kimhi A. (2007): Regression Based inequality Decomposition: A Critical Review and Application to Farm Household Income Data. Discussion Paper No.16.07. The Department of Agricultural Economics and Management. The Hebrew University of Jerusalem.

Morduch J. and Sicular T. (2002): Rethinking Inequality Decomposition, with Evidence from Rural China. The Economic Journal 112:93-106.

ODPM (2004): The English Indices of Deprivation, revised 2004. Office of the Deputy Prime Minister Publication

OECD (2007) Health Data 2007. Organisation for Economic Cooperation and Development. Paris.

Shorrocks, A.F. (1982) Inequality decomposition by factor components. Econometrica, 50: 193 – 211

Shorrocks, A.F (1983): The impact of income components on the distribution of family incomes. Quarterly Journal of Economics, 98(2), 311-326

Sundquist J., Johansson S.E. (1998): The influence of socioeconomic status, ethnicity and lifestyle on body mass index in longitudinal study. International Journal of Epidemiology, 27, 57-63.

Van Doorslaer, E., and X. Koolman (2002). Explaining the differences in income related health inequalities across European countries, Health Economics, 13: 609-628.

Wagstaff, A., E. Van Doorslaer, and N. Watanabe (2004). On decomposing the causes of health sector inequalities, with an application to malnutrition inequalities in Vietnam. Journal of Econometrics, 112(1): 219-27

WHO (1998). Obesity: Preventing and managing the global epidemic. Report of WHO consultation on obesity. Geneva: World Health Organisation.

Zagorsky J.L (2005): Health and Wealth. The late 20th century obesity epidemic in the US. Economic and Human Biology, 3 296 – 313.

Zhang Q., Wang Y. (2004): Socioeconomic inequality and obesity in the United States: do gender, age and ethnicity matter? Social Science and Medicine, 58 1171-1180

Appendix I

Decomposition conditions

Six conditions (based on Shorrocks 1982, 1983 and Fields 2004) under which the s-weights and p-weights are the same for any measure of dispersion that is continuous, symmetric, and take value zero when all Y are identical. (namely the Gini Coefficient, Theil index, and Atkinson index).

1. The inequality measure I(Y) is to be divided into K components, one for each regressors, denoted by $S_k = S_k(Y^1,...,Y^k;K)$
2. Each S_k is continuous in Y_k.
3. The amount of inequality accounted for by any one factor S_k does not depend on how the other factors are grouped i.e. $S_k(Y^1,...,Y^k;K) = S_k(Y^k,Y)$
4. The contribution of S_k sum to the overall amount of inequality
5. If P is any n * n permutation matrix, then $S(Y^kP,YP) = S(Y^k,Y)$. Further if all individuals i have the same value for the k'th factor, then the share of inequality accounted for by that factor is $S_k(\mu_k e,Y) = 0$ for μ_k.
6. Suppose the distribution of Y1 is only a permutation of Y2. If Y1 and Y2 are the only two components in the decomposition, then they should receive the same share in the decomposition.

GE, *A* and the Gini coefficient

Based on information theory the Generalized Entropy measures, $GE(\alpha)$, are computed by (Cowell, 1995):

$$GE(\alpha) = \frac{1}{(\alpha^2 - \alpha)} \left[\left[\sum_{i=1}^{n} f_i \left(\frac{y_i}{\bar{y}} \right)^{\alpha} \right] - 1 \right], \ \alpha \neq 0{,}1 \qquad \text{[A1]}$$

y_i is then BMI of individual i, and there are i=1,2,…n individuals in total. Each individual has a population weight denoted by w_i. Let $f_i = w_i/N$, where $N = \sum_{i=1}^{n} w_i$. When the data are unweighted then w_i = 1 and $N = n$. The mean BMI is given by $\bar{y} = \sum_{i=1}^{n} f_i y_i$. In the present context the parameter α is the weight given to distances between BMI levels between individuals at different parts of the distribution. It can take any real value. For lower values of α, $GE(\alpha)$ is more sensitive to differences at the lower end of the distribution, while at higher values $GE(\alpha)$ is more sensitive to differences at the upper end. $GE(\alpha)$ is commonly estimated for values of α = 0, 1, 2. α = 0 gives more weight to distances between BMI in the lower tail, α = 1 applies equal weight across the distribution and α = 2 gives greater weight to distances in the upper tail. GE(0) and GE(1) are often referred to as the mean logarithmic deviation and the Theil index, respectively. GE(2) is half the squared coefficient of variation. GE(0) and GE(1) are special cases of equation [1], given by

$$GE(0) = \sum_{i=1}^{n} f_i \left[\ln \left(\frac{\bar{y}}{y_i} \right) \right] \qquad \text{[A2]}$$

and

$$GE(1) = \sum_{i=1}^{n} f_i \left(\frac{y_i}{\bar{y}} \right) \left[\ln \left(\frac{y_i}{\bar{y}} \right) \right] \qquad \text{[A3]}$$

, respectively. The value of $GE(\alpha)$ ranges from zero to infinity. A value of zero indicates an equal BMI distribution, and higher values represent higher levels of inequality.

An alternative set of inequality measures is based on the social welfare function approach, quantifying how far actual mean social utility from a given distribution of utilisation deviates from potential mean social utility if BMI were distributed equally across individuals. To compute the Atkinson inequality indices we first define the equally distributed equivalent BMI y_{ε} as (Atkinson, 1972, Cowell, 1995):

$$y_{\varepsilon} = \left[\sum_{i=1}^{n} f_i y_i^{(1-\varepsilon)} \right]^{\left[\frac{1}{(1-\varepsilon)}\right]}, \ \varepsilon > 0, \varepsilon \neq 1 \qquad [A4]$$

ε is an inequality aversion parameter with values ranging from zero to infinity. It is a measure of the relative sensitivity to transfers of BMI at different BMI levels in the distribution. A special case arises when $\varepsilon = 1$, in which case:

$$y_{\varepsilon} = \left[\sum_{i=1}^{n} f_i \ln(y_i) \right], \ \varepsilon = 1 \qquad [A5]$$

The Atkinson indices, $A(\varepsilon)$ are defined by:

$$A(\varepsilon) = 1 - \left(\frac{y_{\varepsilon}}{\bar{y}} \right) \qquad [A6]$$

As ε rises more weight is attached to transfers at the lower end of the distribution and less weight to transfers at the top. The limiting case at the upper end is $\varepsilon \to \infty$, which results in the function A = min y_i, implying that BMI transfers only to the very lowest BMI values count. At the other extreme ε = 0, which ranks distributions solely according to total BMI. $A(\varepsilon)$ is commonly estimated for values of ε = 0.5, 1, 2. $A(\varepsilon)$ may take any value between zero and +1 where higher values represent greater inequality.

The Gini coefficient is computed as:

$$\text{Gini} = \frac{2}{\bar{y}} \sum_{i=1}^{n} f_i y_i Q_i - 1 \qquad [A7]$$

where individuals are ranked in ascending order of BMI y and Q_i is then computed as the weighted relative fractional rank of individual i. For a given individual $i = m$ this is given by:

$$Q_m = \frac{f_m}{2} + \sum_{i=1}^{m-1} f_i \qquad [A8]$$

where $f_0 = 0$.

The Gini coefficient may take a value between zero and +1. A value of zero indicates an equal distribution of BMI across population. A value of +1 means that all body mass is concentrated at the top end of the BMI distribution.

Appendix II

Table A1. Regression results and decomposition results for

Dependent variable BMI	*Coefficients.*	*t values*	*s weights*	*p weights*
Income squared	0.008242	1.64	0.002124	0.013643
Income	-1.1E-05	-1.3	-0.00158	-0.01012
Age	0.572887	5.94	0.424451	2.72631
Age squared	-0.00855	-4.17	-0.44717	-2.87221
Age cube	3.68E-05	2.73	0.106833	0.686204
No Qualification	0.778693	3.04	0.00436	0.028005
GCSE	0.304032	1.39	-0.00093	-0.00595
A Level	0.549753	2.18	-0.00064	-0.00412
Other qualification	1.267575	0.97	0.000266	0.001709
Black	-0.46952	-1.19	0.001172	0.007528
Asian	0.752169	1.01	0.00002	0.000128
Single	-0.76474	-2.77	0.019181	0.123202
Separated	-1.155	-2.16	-0.00016	-0.00102
Divorced	-0.77478	-2.2	-0.00059	-0.0038
Widowed	0.320131	0.61	0.00003	0.000193
ILO unemployed	-0.13757	-0.37	0.000636	0.004085
Retired	0.019561	0.05	0.000036	0.000231
Economically inactive	0.166796	0.56	-0.00061	-0.00394
Rent	0.654714	2.44	-0.00331	-0.02128
Mortgage	0.731288	3.13	0.002755	0.017696
Social Class 2	0.820728	2.93	0.004988	0.032039
Social Class 3 and 4	0.738977	2.59	0.004503	0.028923
Social Class 5 and 6	0.190098	0.6	-0.00069	-0.00443
One child in Household	-0.37809	-1.42	0.001213	0.007791
Two children in Household	-0.0082	-0.02	-6.80E-06	-4.4E-05
Three children in Household	-0.37738	-0.67	0.000224	0.001439
Four children in Household	0.411286	0.27	0.000028	0.00018
Five Children in Household	-1.98566	-0.37	0.000016	0.000103
Two persons living in Household	0.665961	2.55	0.008729	0.056068
Three persons living in Household	-0.07361	-0.27	0.00011	0.000707
Four persons living in Household	-0.39267	-1.35	0.001821	0.011697
Five or more persons living in Household	-0.13661	-0.26	0.00027	0.001734
Second Degree of Area Deprivation	0.242522	0.99	0.000268	0.001721
Third of higher Degree of Area Deprivation	0.517575	2.24	0.000467	0.003
One infant in Household	0.69624	1.72	0.000486	0.003122
Two infant in Household	1.18718	0.63	0.000091	0.000585
Strategic Health Authority 2	-0.72622	-1.29	0.00072	0.004625

Strategic Health Authority 3	0.62757	1.14	0.000935	0.006006
Strategic Health Authority 4	-1.09858	-1.9	0.002461	0.015807
Strategic Health Authority 5	-0.60092	-0.92	0.000887	0.005697
Strategic Health Authority 6	-0.12664	-0.22	0.000038	0.000244
Strategic Health Authority 7	-1.32515	-1.88	0.000787	0.005055
Strategic Health Authority 8	-1.94901	-2.93	0.002728	0.017522
Strategic Health Authority 9	0.245571	0.43	-5.9E-05	-0.00038
Strategic Health Authority 10	0.3005	0.47	0.000138	0.000886
Strategic Health Authority 11	0.233667	0.44	0.000084	0.00054
Strategic Health Authority 12	-0.53976	-1.03	0.000334	0.002145
Strategic Health Authority 13	0.178194	0.33	0.000154	0.000989
Strategic Health Authority 14	0.055427	0.11	0.00004	0.000257
Strategic Health Authority 15	-0.32033	-0.61	0.000171	0.001098
Strategic Health Authority 16	0.113567	0.22	-3.9E-05	-0.00025
Strategic Health Authority 17	-0.35465	-0.62	0.000213	0.001368
Strategic Health Authority 18	0.022333	0.04	0.000012	7.71E-05
Strategic Health Authority 19	-0.43409	-0.9	0.000335	0.002152
Strategic Health Authority 20	0.240522	0.48	0.000208	0.001336
Strategic Health Authority 21	0.196549	0.35	0.000271	0.001741
Strategic Health Authority 22	-0.14687	-0.25	0.000011	7.07E-05
Strategic Health Authority 23	-1.15079	-1.82	0.001701	0.010926
Strategic Health Authority 24	0.483578	1.02	0.001123	0.007213
Strategic Health Authority 25	-0.11571	-0.21	-1.4E-05	-9E-05
Strategic Health Authority 26	-0.34817	-0.6	0.00004	0.000257
Strategic Health Authority 27	0.381556	0.73	0.000431	0.002768
Strategic Health Authority 28	0.782265	1.42	0.001246	0.008003
Interview in January	0.848446	2.3	0.000979	0.006288
Interview in February	0.336726	0.85	-6.30E-06	-4E-05
Interview in March	0.31566	0.85	-4.6E-05	-0.0003
Interview in April	0.273386	0.76	-0.00034	-0.00215
Interview in May	0.391229	1.02	-0.00016	-0.00105
Interview in July	0.105457	0.31	-0.00011	-0.00071
Interview in August	0.363484	0.99	0.000194	0.001246
Interview in September	0.355703	0.96	-0.00033	-0.00214
Interview in October	0.675281	1.77	0.000893	0.005736
Interview in November	0.533374	1.57	0.000282	0.001811
Interview in December	0.96022	2.29	0.001924	0.012358
Daily portions of fruit	0.098785	1.6	0.001803	0.011581
Daily portions of fruit squared	-0.00558	-1.26	-0.00034	-0.0022
Household owns one car	0.882912	3.35	0.002685	0.017246
Household owns two cars	0.953001	3.21	0.003856	0.024768
Household owns three or more cars	1.632762	4.57	-0.00076	-0.00488
Missing values / ethnicity	3.352022	2.09	0.000381	0.002447
Missing values / employment status	-0.48753	-0.32	0.0002	0.001285
Missing values / tenure	2.772095	1.91	0.000093	0.000597
Missing values / car ownership	1.713546	0.52	0.000143	0.000919
Constant	13.47858	8.6		
R^2	0.1747			
N	2923			

Table A2. Regression results and decomposition results for women.

Dependent variable BMI	*Coefficient*	*t value*	*s weights*	*p weights*
Income squared	0.012815	1.61	-0.00401	-0.03801
Income	-2.8E-05	-2.3	0.011376	0.107899
Age	0.37627	3.05	0.195206	1.851487
Age squared	-0.00408	-1.57	-0.17585	-1.66791
Age cube	7.40E-06	0.44	0.021683	0.205659
No Qualification	0.934652	2.82	0.009786	0.092818
GCSE	0.294784	1.02	-0.00054	-0.00514
A Level	0.66361	1.94	-0.00134	-0.01272
Other Qualification	-0.16933	0.86	0.000329	0.00312
Black	0.845091	-0.31	0.000215	0.002039
Asian	0.553328	1.12	-0.00005	-0.00047
Single	-0.36943	-1.11	0.004464	0.04234
Separated	-0.34201	-0.54	-5.5E-05	-0.00052
Divorced	-0.34847	-0.96	-0.00055	-0.00518
Widowed	0.238409	0.52	0.00065	0.006165
ILO unemployed	-0.61624	-1.2	0.001624	0.015403
Retired	0.891315	2.22	0.005393	0.051151
Economically inactive	0.406282	1.47	0.000634	0.006013
Rent	0.723056	2.29	0.002205	0.020914
Mortgage	0.633112	2.23	-0.00314	-0.02981
Social Class 2	0.754065	1.93	-0.00186	-0.01762
Social Class 3 and 4	0.584875	1.57	-0.0002	-0.0019
Social Class 5 and 6	0.965197	2.43	0.00724	0.06867
One child in Household	0.321812	1.06	0.000396	0.003756
Two children in Household	-0.99555	-2.56	4.12E-03	0.039115
Three children in Household	0.062936	0.09	0.000019	0.00018
Four Children in Household	-1.81575	-1.14	0.000249	0.002362
Five or more children in Household	2.807213	0.41	0.000098	0.00093
Two persons living in Household	0.045467	0.14	0.000091	0.000863
Three persons living in Household	0.742447	2.17	0.00154	0.014607
Four persons living in Household	0.473464	1.31	-0.00148	-0.014
Five or more persons living in Household	1.434716	2.15	0.000161	0.001527
Second Degree of Area Deprivation	0.493473	1.67	0.000064	0.000607
Third of higher Degree of Area Deprivation	0.574448	2.05	0.002278	0.021606
One infant in Household	0.519431	1.18	-0.00018	-0.00175
Two infant in Household	-0.58738	-0.27	0.000059	0.00056
Strategic Health Authority 2	-1.10898	-1.65	0.000649	0.006156
Strategic Health Authority 3	-0.51454	-0.78	3.20E-06	3.04E-05
Strategic Health Authority 4	-2.89104	-4.16	0.005852	0.055505
Strategic Health Authority 5	-2.91828	-3.73	0.00438	0.041543
Strategic Health Authority 6	-1.26214	-1.76	0.000708	0.006715
Strategic Health Authority 7	-0.97488	-1.1	0.000332	0.003149
Strategic Health Authority 8	-2.75308	-3.49	0.004088	0.038774
Strategic Health Authority 9	-0.41778	-0.6	-0.00042	-0.00402
Strategic Health Authority 10	-0.40729	-0.53	-0.00022	-0.00206

Strategic Health Authority 11	-0.4391	-0.69	-0.00018	-0.00168
Strategic Health Authority 12	-0.81621	-1.26	-9.4E-05	-0.00089
Strategic Health Authority 13	-0.33183	-0.51	-0.00019	-0.00182
Strategic Health Authority 14	-1.17079	-1.92	0.000505	0.00479
Strategic Health Authority 15	0.139973	0.23	0.000163	0.001546
Strategic Health Authority 16	-0.24793	-0.38	-1.60E-06	-1.5E-05
Strategic Health Authority 17	-1.16608	-1.69	0.000396	0.003756
Strategic Health Authority 18	-1.74382	-2.58	0.001955	0.018543
Strategic Health Authority 19	-0.98072	-1.65	0.001184	0.01123
Strategic Health Authority 20	-0.25595	-0.41	-0.00011	-0.00107
Strategic Health Authority 21	-0.65849	-0.94	-0.00011	-0.00104
Strategic Health Authority 22	-0.43505	-0.6	-7.3E-05	-0.00069
Strategic Health Authority 23	-0.72908	-1.01	0.000088	0.000835
Strategic Health Authority 24	-0.14458	-0.24	-0.00011	-0.00108
Strategic Health Authority 25	0.518467	0.73	0.000612	0.005805
Strategic Health Authority 26	-0.50583	-0.77	0.000066	0.000626
Strategic Health Authority 27	-0.45535	-0.72	-0.00039	-0.00371
Strategic Health Authority 28	0.75124	1.11	0.00104	0.009864
Interview in January	-0.2742	-0.63	0.000157	0.001489
Interview in February	0.182002	0.38	1.42E-04	0.001347
Interview in March	-0.2274	-0.5	-2.7E-05	-0.00026
Interview in April	-0.03854	-0.09	-5.90E-06	-5.6E-05
Interview in May	-0.2682	-0.59	0.000053	0.000503
Interview in July	-0.64506	-1.55	0.001179	0.011183
Interview in August	-0.16115	-0.37	0.00007	0.000664
Interview in September	0.025597	0.06	0.000014	0.000133
Interview in October	0.191801	0.43	0.00023	0.002182
Interview in November	0.001294	0	5.20E-07	4.93E-06
Interview in December	-0.43229	-0.83	0.000242	0.002295
Daily portions of fruit	-0.05013	-0.62	0.000331	0.003139
Daily portions of fruit squared	0.006034	0.97	-0.0003	-0.00288
Household owns one car	0.008069	0.03	0.000026	0.000247
Household owns two cars	-0.36635	-1.05	0.001622	0.015384
Household owns three or more cars	-0.32814	-0.73	0.000473	0.004486
Missing values / ethnicity	2.801036	1.07	-4.8E-05	-0.00046
Missing values / qualification	-8.85035	-1.4	0.000446	0.00423
Missing values / employment status	0.055417	0.02	-4.00E-06	-3.8E-05
Missing values / tenure	0.427989	0.23	-4.7E-05	-0.00045
Missing values / Car ownership	-4.60645	-0.79	0.000137	0.001299
Constant	17.74099	9.01		
R^2	0.1138			
N	3388			