HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 15/22

The Income-Health Relationship "Beyond the Mean": New Evidence from Biomarkers

Vincenzo Carrieri & Andrew M. Jones

September 2015

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

The Income-Health Relationship "Beyond the Mean": New Evidence from Biomarkers

Vincenzo Carrieri*
Università di Salerno
HEDG, University of York

Andrew M. Jones
University of York
Monash University
University of Bergen

Abstract

This paper offers new evidence on the income-health relationship by analyzing the income gradient across the full distribution of four blood-based biomarkers: cholesterol, fibrinogen, glycated haemoglobin and ferritin. We use an unconditional quantile approach based on recentered influence function (RIF) regressions and apply an Oaxaca-Blinder decomposition at various quantiles of biomarker distributions to explain gender differentials in biomarkers. Using ten waves of the Health Survey for England (from 2003 to 2012) we find a non-linear relationship between income and biomarkers and a higher income gradient at the highest quantiles of the biomarker distributions. Moreover, we find that there is an important heterogeneity in the association of health to income across genders which varies significantly along the biomarker distribution and accounts for a substantial percentage of the gender differentials in observed health.

Keywords: biomarkers; unconditional quantile regression; decomposition analysis; health inequalities.

JEL codes: C1, C5, I14.

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

Introduction

The positive association between income and health is a well-established finding in the health economics literature. This relationship has been found across age groups, in many countries analyzed and for a variety of health measures, including self-rated health (Mackenbach et al., 2005; Ettner, 1996), functional limitations (Ettner, 1996), anthropometric measures (Wagstaff, Van Doorslaer and Watanabe, 2001) and mortality (Cutler, Deaton and Lleras-Muney, 2006). The influence of economic conditions on health is also the basis of the research on "income-related health inequalities", one of the most prominent fields in the health economics literature (see e.g., Kakwani, Wagstaff and van Doorslaer, 1997; van Doorslaer and Jones, 2003). Studies of the influence of economic conditions on health typically measure the effect of the former on the conditional mean of the health status variable. Unfortunately, analysis based solely on the mean misses potentially important information in other parts of the distribution (Bitler et al., 2006). This is especially relevant to the income-health relationship, where clinical concern is focused on the tail of the distributions and where evaluating the income gradient at different points of the distribution of health status and decomposing income-related inequalities in health could be beneficial (Jones and Lopez, 2006).

This lack of evidence is likely due to two factors. On one hand, health information is often unavailable on a continuous scale in standard health or social surveys. For instance, self-assessed health and functional limitations are collected on an ordinal scale while mortality is a dichotomous indicator (by nature). "Beyond the mean analysis" is obviously less attractive in these cases. On the other hand, the literature in econometrics has developed techniques going "beyond the mean" only recently (see Fortin et al., 2011 for a review). This is because, unlike average estimation framework, the estimates on the entire distribution of the dependent variables, i.e. the quantile regression, cannot be easily used to estimate the impact of a covariate on the corresponding unconditional quantile of the dependent variable (Firpo, Fortin and Lemieux, 2009). OLS regressions provide consistent estimates of the impact of an explanatory variable, X, on the population unconditional mean of an outcome variable, Y, because the conditional mean, $E[Y|X]$, averages up to the unconditional mean, $E[Y]$, due to the law of iterated expectations. As a consequence, a linear model for conditional means $E[Y|X] = X\beta$ implies that $E[Y] = E[X]\beta$ and OLS estimates of β also indicate what is the impact of X on the population average of Y (Firpo, Fortin and Lemiux, 2009). When the attention shifts towards the entire distribution, the situation is more complicated because conditional quantiles do not average up to their population counterparts, ie: $q_y(\tau) \neq E[q_{y|X}(\tau)]$. The analysis "beyond the mean" of the unconditional distribution of the dependent variable is then more challenging.

In this paper, we use a distributional method proposed in the recent literature, the recentered influence function approach (RIF) of Firpo, Fortin and Lemieux (2009) to estimate the income gradient for a continuous measure of objective health status: blood-based biomarkers. Biomarkers are characteristics that are 'objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacologic responses

to a therapeutic intervention'[1]. They are measured on a continuous scale associated with an increasing or decreasing risk (depending on the biomarker) of a disease state and they are often highly correlated with mortality (Bixby and Dow, 2012; Sattar et al., 2009; Gruenewald et al., 2006). We consider four blood-based biomarkers available in ten waves of the Health Survey for England (2003-2012) associated with the most relevant diseases in all Western countries: cholesterol, glycated haemoglobin, fibrinogen and ferritin. Cholesterol measures "fat in the blood" and it is associated with a higher risk of heart disease; glycated haemoglobin is a biomarker for diabetes; ferritin is a biomarker for poor nutrition and is associated with other important diseases such as liver diseases; fibrinogen is a haemostatic marker associated with many inflammatory diseases including cardiovascular and liver diseases.

Biomarker data have also been recently used to analyze the effect of socioeconomic position on the conditional mean of the biomarker score. Using biomarker data for diabetes, hypertension and cardiovascular diseases, Banks et al. (2006) found that English residents have on average better health than US residents. Juerges, Kruk and Reinhold (2013) found a positive relationship between schooling and biomarkers of cardiovascular diseases (fibrinogen and C-reactive protein). Muennig, Sohler and Mahato (2007) look at differences between socioeconomic groups in C-reactive protein and cholesterol homocysteine, associated with cardiovascular diseases. They found a positive effect of income and education on "good cholesterol" and a slightly significant effect on fibrinogen. Ploubidis et al. (2014) found a negative impact of early life socioeconomic position on fibrinogen levels later in life. Dowd and Goldman (2006) tested the influence of stress biomarkers on the relationship between socioeconomic status and health. They found that chronic stress is actually not very different across socioeconomic groups. A key advantage from using biomarker data in all these analyses is having a measure of health which is free of reporting bias. This is particularly important given the intense debate around the extent of socioeconomic-related reporting bias.[2]

All these studies corroborate the idea that social position and economic conditions contribute to shape objective health, i.e. biomarkers scores. Our paper goes beyond this literature in two ways. Firstly, we use the unconditional quantile regression approach developed by Firpo, Fortin and Lemieux (2009) to estimate the impact of a marginal change in income on the entire distribution of biomarkers. The unconditional quantile method consists of running a regression of the (recentered) influence function of the unconditional quantile on the explanatory variables. The method works by providing a linear approximation to a non-linear functional of the distribution. This allows one to apply the law of iterated expectations to the distributional statistics of interest (i.e. the quantiles) and thus to compute approximate partial effects of a single covariate on the functional being approximated. In our setting, this method allows us to properly assess the impact of income

[1] This definition is given by National Institute of Health Biomarkers Definitions Working Group (2001).

[2] Empirical literature investigating the extent of reporting heterogeneity in health between socioeconomic groups has found contrasting results. Some papers (i.e. Sen (2002) for India) show the existence of a large reporting heterogeneity. Some others, (i.e. Hernández-Quevedo, Jones and Rice, (2004) for the United Kingdom) do not find any heterogeneity while some others (i.e Bago d'Uva et al. (2008) for China, Indonesia and India) found only a small reporting heterogeneity and not in all countries analyzed.

at different points of the unconditional distribution of biomarkers. This is desirable in order to check for non-linearities in the relationship between income and health and to assess the role of income at extreme biomarker levels which often indicate the presence or the risk of severe diseases that are associated with high costs for the health system.

Secondly, we apply an Oaxaca-Blinder decomposition at various quantiles of the biomarker distributions to analyse gender differentials in biomarkers and to measure the contribution of income (and other covariates) to these differentials. Oaxaca-Blinder decomposition assesses to what extent gender differentials in biomarkers are explained by compositional differences, ie. differences in observed covariates, or by differences in elasticity of health with respect to income and other factors. Under the assumption of the *invariance of conditional distribution* (see Fortin, Firpo and Lemieux 2011), discussed in more detail in section 3, this might also identify the causal effect of gender on the biomarkers. More generally, decomposition analysis is important for drawing policy implications because it helps to deliniate the role of health policies, that operate on the health-gradient, from the role of fiscal policy, that operates on compositional differences (see Van Doorslaer and Koolman, 2004 for a discussion of the policy implications of decomposition analysis). Jones and Lopez (2006) have already found that there is a great heterogeneity in the association of health with explanatory variables across genders and they have shown that this has important consequences on the measurement of income-related inequalities in health. Our paper contributes to this literature by performing the decomposition analysis of gender differentials in health along the entire distribution of health status: this permits an assessment of gender-related differences in income-health gradient across the entire distribution of health status.

We find a non-linear relationship between income and health and a strong gradient with respect to income at the highest quantiles of the biomarker distributions. In some cases, i.e. cholesterol, the income gradient is found only at high quantiles, while analysis "at the mean" leads to misleading conclusions (ie. a positive relationship between income and cholesterol). This makes the analysis on the entire distribution especially relevant in such cases. Secondly, we find that there is important heterogeneity in the association of health to income across genders which varies significantly along the biomarker distribution and accounts for a substantial percentage of the gender differentials in observed health.

The rest of the paper is organized as follows. The next section presents data and descriptive statistics. Section 3 discusses the empirical methodology. Section 4 presents the results. The final section summarizes and concludes.

2. Data

Our data come from the Health Survey for England (HSE). HSE is an annual health interview survey of around 15,000 to 20,000 respondents in England conducted by the National Centre for Social Research (separate surveys are available for Scotland and Wales).

The survey started in 1991 and has been carried out annually since then. HSE includes adults aged 16 and over, and since 1995 has also included children aged 2-15. From 2001 onwards, the survey covers all ages, but certain age groups are asked questions on selected topics only. An interview with each eligible person in the household is followed by a nurse visit for those who agree to take part. The interview includes a set of core questions, asked each year on general health and psycho-social indicators, smoking, alcohol, demographic and socio-economic indicators, questions about use of health services and prescribed medicines. Biomarkers are collected during nurse visits and include not only blood samples but also anthropometric measurements, blood pressure measurements, and saliva samples. During the nurse visits, the nurse asks the respondent for permission to carry out various types of measurements. Respondents are informed about the purpose of each test and the value of each test for the monitoring of various diseases. For instance, for the cholesterol test, the nurse informs participants that "high levels are associated with blood clots, heart attack and stroke". The delivery of information is useful in order to increase compliance and establish a good working relationship.

The most popular blood-based biomarkers, which are analyzed in this paper, have been collected since 2002 in HSE. More precisely, Cholesterol and glycated hemoglobin were collected since 2003 to 2012 every year. Fibrinogen was collected since 2003 to 2006 and in 2009. Ferritin was collected in 2002, from 2004 to 2006 and in 2009. Other potentially relevant biomarkers (ie. tryglicerides, C-reactive protein) are collected sporadically and they are not included in this analysis. We do not make any statistical transformation of the blood-based biomarkers sample and use the valid (ie. blood sample properly collected and successfully processed) biomarker measurements in each wave. Thus, we can use 30,770 non-missing observations for the analysis of cholesterol spanning over the period 2003-2012, 34,831 for the analysis of glycated haemoglobin over the period 2003-2012, 15,530 for the analysis of fibrinogen from 2003 to 2006 and in 2009 and 14,188 observations for ferritin available in 2002, 2004, 2005, 2006 and 2009. An implicit age stratification comes from the age restriction used by HSE for the blood sample collection. For almost all the waves, blood samples are collected from individuals aged 16+. In a few waves a different age restriction is employed. In the 2002 only individuals aged 24 or less are included; in the 2004 individuals aged 11+ are included, while in 2005 only individuals aged 65+ are analyzed. A careful control for demographics is employed in all our analysis and sample weights are used in all regressions to take account of the survey design.

2.1 Variables and descriptive statistics

In what follows, we provide a description of the variables used in our analysis. Firstly, we describe the blood-biomarker variables giving some detail on their unit of measurement, the clinical cutpoints (when available) and the use of biomarker values for diagnosis of a disease state. Later, we describe the income variable and other controls employed in the regression. A complete list of the variables along with some descriptive statistics is then presented in Tables 1 and 2.

We examine four blood-based biomarkers: Total Cholesterol, Glycated Haemoglobin, Fibrinogen and Ferritin. As discussed in the introduction, these markers are highly predictive of the most relevant chronic diseases in Western Countries such as England. Total cholesterol (TC) is measured in units called millimoles per litre of blood, (mmol/L). The English government recommends that total Cholesterol should be equal or less than 4 mmol/L among individuals at high risk of cardiovascular disease (CVD) (ie. obese, with an history of CVD, etc.) and equal or less than 5mmol/L or less for healthy individuals. Values above these thresholds indicate a higher risk of CVD.

Glycated haemoglobin (HbA1c) is a measure of the level of sugar in the blood over the previous 8 to 12 weeks before measurement. It is the proportion of haemoglobin proteins that have been bound by glucose. HbA1c can be expressed as a percentage or as a value in mmol/mol. HbA1c is measured in percentage in all waves of the HSE. HbA1c values $\geq$ 6.5% indicates diagnosis of diabetes, while values between 5.7% and 6.4% indicate pre-diabetes risk (ADA, 2010; WHO, 2011a).

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

Levels of Ferritin reflect the size of the body's stock of iron and therefore they are indicative of anaemia. A low ferritin level is predictive of uncomplicated iron deficiency anaemia, caused, for instance, by poor nutrition. However, high ferritin levels suggest excess body iron, which is also problematic for health because it is generally associated with important diseases such as liver diseases. WHO (2011b) suggests some cut-points: ferritin levels below $\leq$ 20 ug/L indicate depletion of iron, while levels $\leq$ 12 indicate complete absence of stored iron. Ferritin levels >300 ug/L may indicate iron overload in men and post-menopausal women and >200 may indicate iron overload in pre-menopausal women.

Our main independent variable is household equivalised income. It includes total income of a household from all sources, after tax and other deductions, divided by the number of household members converted into equivalised adults. In order to take into account the fact that income changes are often multiplicative in the real world (ie. a 5% raise in wages), we take the logarithms of equivalised income in all regressions. This allows a better interpretation of the income-health relationship.

As control variables, we include six age group variables (11-18, 18-34, 35-44, 45-64, 65-74, 75+) for each gender and seven dummies for educational status. Education is measured according to the following categories: Degree or National vocation qualification (NVQ) 4 or 5; Higher education below degree; NVQ 3 or General Certificate of Education (GCE) Advanced Level; NVQ 2 or GCE ordinary level; NVQ1 or Certificate of Secondary Education (CSE); Other qualifications from outside England; No qualification. Omitted categories in our analysis are males aged 11-18 and individuals with no qualification.

Table 1 shows some descriptive statistics for the biomarkers. We find that average biomarker values in our sample fall essentially within normal ranges, but with some distinctions. In particular, cholesterol values are just a bit higher than the cutpoint of 5 while fibrinogen average scores are a bit higher than the normal cutpoint of 3. Moreover, Table 1 depicts a higher dispersion around the average cholesterol and Ferritin values, while other biomarkers values are less dispersed around the mean.

Table 2 shows descriptive statistics for the independent variables. We observe a high share of individuals aged 45-64 and a higher share of elderly women (over 65), consistent with the well-known gender differences in life expectancy. With respect to education, we observe that around 23% of individuals in our sample have a degree, and a similar share of them have a NVQ 2 or GCE. Also the share of individuals without formal education is substantial (around 21%). This includes mostly individuals belonging to older cohorts without any formal education (around 20% and an average age of 60) and students receiving compulsory education or attending a school at the time in which the interview was conducted (approximately 1% of our sample and an average age of 15).

3. Empirical Methodology

Our empirical analysis is based on the Recentered Influence Function (RIF) method of Firpo, Fortin and Lemiux (2009). We use the RIF method to estimate the relationship between income and biomarkers and we use the RIF regression as a basis for Oaxaca-Blinder decomposition of gender differentials in England.

As discussed in the introduction, the key advantage of the RIF approach is that it allows us to analyze the relationship between income and the *unconditional* distribution of biomarkers, and to analyze and decompose differences in the *unconditional* distribution of biomarkers across genders. This possibility is essentially given by the fact that RIF method works by providing a linear approximation of the unconditional quantiles of the dependent variable. The law of iterated expectations can be applied to the quantile being approximated and used to estimate the marginal effect of a covariate through a simple regression of a function of the outcome variable, the Recentered Influence Function, on the covariates X.

In our setting, the RIF of biomarkers is estimated directly from the data by first computing the sample quantile q and then estimating the density of the distribution of biomarkers at that quantile using kernel density methods. Then, for a given observed quantile q_τ, a RIF is generated which can take one of two values depending upon whether or not the observation's value of the outcome variable is less than or equal to the observed quantile:

$$RIF\ (Bio; q_\tau) = q_\tau + \frac{\tau - 1[Bio \leq q_\tau]}{f_{Bio}(q_\tau)} \quad (1)$$

Where q_τ is the observed sample quantile, $1[Bio \leq q_\tau]$ is an indicator variable equal to one if the observation's value of the biomarker is less than or equal to the observed quantile and zero otherwise. $f_{Bio}(q_\tau)$ is the estimated kernel density of the biomarker at the τ_{th} quantile.

The RIF defined in equation (1) is then used as a dependent variable in a OLS regression on the covariates X, as defined in section 2.1. In practice, this amounts to estimate a rescaled linear probability model (Jones, Lomas and Rice, 2015). Indeed, the unconditional quantile of the biomarker ,q_τ, may be obtained as follows:

$$q_\tau = E_x\left[E\left[\widehat{RIF}(Bio; q_\tau)|X\right]\right] \quad (2)$$

Where $\widehat{RIF}(Bio; q_\tau)|X$ is the estimate of RIF as defined in equation (1) conditional on covariates X. Thanks to this linear approximation, it is now possible to apply the the law of iterated expectations. Thus, q_τ can be written as :

$$q_\tau = E[X]\widehat{\delta_\tau}$$

where $\widehat{\delta_\tau}$ is the coefficient of the unconditional quantile regression. This linearization allows estimation of the marginal effect of a change in distribution of covariates X (including income) on the *unconditional* quantile of biomarkers, measured by the parameter $\widehat{\delta_\tau}$. In our model, as well as the covariates X presented in Table 2, we also include year fixed effects, to pick up time variation in biomarker levels.

To analyze gender differentials in biomarkers, we use Oaxaca-Blinder (Blinder 1973; Oaxaca 1973) (OB) decomposition method using the RIF regression in equation (2) as a basis for the decomposition. A similar logic to the OB decomposition at the mean applies also here (see Fortin, Lemieux, Firpo, 2011 for a review). Formally, differences in estimated biomarkers levels between males (M) and females (F) year at each quantile can be decomposed as follows:

$$\Delta^\tau_{Bio} = \widehat{[RIF}(Bio_M, q_{M\tau})] - \widehat{[RIF}(Bio_F, q_{F\tau})]$$

$$\Delta^\tau_{Bio} = (\bar{X}_M - \bar{X}_F)\delta_F + \bar{X}_F(\delta_M - \delta_F) \quad (3)$$

where $\bar{X}_M$ and $\bar{X}_F$ represent the sample means of covariates X for the subsample of males and females, and δ_M and δ_F represent the coefficients of the unconditional quantile regression as in equation (2) for the subsample of males and females, respectively.

The first term in equation (3) is the part of differential in biomarkers that is "explained" by differences in observed covariates between the subsample of males and females. This is often called as a "composition effect". Differences in covariates across genders are weighted by the coefficients of the unconditional quantile regression from a model estimated on the subsample of females (δ_F). The decomposition is thus formulated from the viewpoint of females as in the original work by Oaxaca (1973). In our application, the choice of the discriminated group is complicated by the fact that women might have some health advantages over men (ie. they have a higher longevity than men, for instance) but they often

earn less than males because of gender discrimination in the labour market. We therefore consider females as the discriminated group. This puts our OB decomposition in line with the traditional discrimination literature, ie. the analysis of gender wage gap (see the discussion in Neumark, 1988 and Jann, 2008 for more details).

The second term in equation (3) measure the "unexplained" part of the differential in biomarkers. This is often called also as "structural" part and it accounts for differences in biomarkers across genders which is due to differences in the impact of the covariates and it also captures all potential effects of differences in unobserved variables.

The explained and unexplained part can be further decomposed into contributions of each covariate at each quantile. In our case, it is particularly useful to derive both the total contribution and the detailed contribution of income to the gender differentials in biomarkers. This allows to understand to what extent differences in biomarkers are driven by differences in earnings between males and females ("composition effect") and/or by differences in the association of health to income across genders (also known as "elasticity effect"). Thanks to the additivity assumption of the OB decomposition, this is possible because the "explained" and "unexplained" part in equation (3) are simply given by the sum of the contribution of individual covariates.

Thus, it is possible to derive the detailed contribution of all covariates (including income) to the "explained part", as follows:

$$(\bar{X}_M - \bar{X}_F)\delta_F = (\bar{X}_{1M} - \bar{X}_{1F})\delta_{1F} + (\bar{X}_{2M} - \bar{X}_{2F})\delta_{2F} + \dots \quad (4)$$

Where $\bar{X}_1$ and $\bar{X}_2$ … are the means of the single covariates and δ_F are the associated coefficients of the unconditional quantile regression estimated on the subsample of females. Similarly, the contributions of each covariate to the "unexplained part" can be obtained as as follows:

$$\bar{X}_F(\delta_M - \delta_F) = \bar{X}_{1F}(\delta_{1M} - \delta_{1F}) + \bar{X}_{2F}(\delta_{2M} - \delta_{2F}) + \cdots \quad (5)$$

To draw inference on the contributions of each covariate to the explained and unexplained part, standard errors are computed using the delta method (See Jann, 2008 for more details).

First to present our results we highlight some interesting connections between the decomposition presented above and the treatment effect literature. In principle, if we consider the gender as a treatment, decomposition presented in equations (3)- (5) might have a causal interpretation. Indeed, the explained part ($(\bar{X}_M - \bar{X}_F)\delta_F$) in equation (3) captures all the compositional differences, ie. in income and in the other factors, between males and females. This part might be thus conceived as the selection bias resulting from confounding factors to be controlled for in the program evaluation literature. Instead, the unexplained part ($\bar{X}_F(\delta_M - \delta_F)$) in equation (3) might be conceived as the treatment effect of gender on health, or, more precisely, as the Population Treatment Effect on the Treated (Fortin, Firpo and Lemiuex 2011). Moreover, thanks to the additivity assumption, the detailed OB decomposition presented in equation (4) and (5) allows us to separate the elasticity from the compositional effect of each covariate on the total differences in biomarkers between males

and females. Thus, for instance, the elasticity effect of income (or of other covariates) in equation (5) can be interpreted as the causal effect of gender on income-health gradient (or other covariates-health relationship). In a recent paper, Sloczynski (2015) also examines the finite-sample performances of the OB unexplained component as an estimator of the PATT and finds that it performs better than many popular methods that have received considerable attention in the treatment effects literature, such as inverse probability weighting estimators, kernel matching, nearest-neighbour matching, etc. (see Sloczynski (2015) for more details) .

However, it is important to note that the interpretation of OB decomposition in causal terms is valid under the assumptions of *ignorability and common support*[3]. These assumptions are also the identifying assumptions of all the estimators mentioned above, belonging to the strand of treatment effect literature which relies on *selection on observables*. When the OB decomposition is performed on the entire outcome distribution, as in our case, these assumptions guarantee the *invariance of conditional distribution*, namely, that the conditional distribution of the biomarker given the control variables X remains invariant under manipulations of the marginal distribution of the X[4]. This permits us to derive a valid counterfactual of the distribution of biomarkers for females using the biomarker-covariates relationship estimated on males (Fortin, Firpo and Lemiuex (2011)).

In our setting, some of these assumptions are rather restrictive while some others seem to be more justified. In particular, the *ignorability assumption* - requiring that the conditional distribution of unobservables to be the same across genders - is restrictive, because important unobservable determinants of health status such as preferences for risk, inter-temporal preferences or variables for which we cannot control for in our dataset (lifestyles, genetic inheritance) might be unevenly distributed across genders. On the other hand, we checked balancing properties of our covariates across gender and found a slight lack of balance for some covariates (ie. demographics and some educational dummies) but a good balance on our key covariate, income[5]. This suggests that *common support* assumption-requiring that the values of the control variables observed for males are also observed among females - might be justified. In view of this, a causal interpretation of the decomposition may be possible.

Despite that, we prefer to not give a causal interpretation of our results essentially because of the nature of our treatment. Gender is not a choice or a manipulable action, as we cannot obviously conceive individuals choosing which group to belong to (see for instance the discussion in Fortin, Firpo and Lemieux, 2011 and in Holland, 1986). Nonetheless, we think that relating the decomposition exercise to the policy evaluation framework is useful for a deeper understanding of the decomposition results shown in the next sections of the paper.

---

[3] Together termed also as *strong ignorability* . Please see Fortin, Firpo and Lemiuex (2011) for a clear discussion on the identifying assumptions of the most popular decomposition methods employed in economic analysis.

[4] An additional assumption, which is somewhat implicit in the OB decomposition, is the simple counterfactual assumption. This implies that other counterfactuals based on hypothetical states of the world, ie. general equilibrium effects, are ruled out. In our case, this means that males are a valid counterfactual for females. Consequently, it means to assume that if females were not penalized, they would exhibit the same health-covariates relationship found among males.

[5] Results are not shown here and they are available upon request.

4. Results

4.1 Income-health relationship

Tables 3-6 shows the results of RIF regressions described in equation (2) for cholesterol, glycated haemoglobin, fibrinogen and ferritin, respectively. Column 1 of each table includes OLS regression at the mean for comparison, while columns 2-5 include results of the RIF regressions at the 25th, 50th, 75th, 90th and 95th percentile of each biomarker, respectively. To make the interpretation of our coefficient of interest easier, we also plot the income coefficient at all points of the biomarker distribution (every 5 percentiles) in Figures 1-4.

Table 3 shows that the relationship between income and cholesterol is somewhat complex and that analysis at the mean misses important information. Indeed, OLS estimates suggest only a positive income gradient (column 1), while RIF regressions indicate that income-cholesterol relationship varies at different points of the cholesterol distribution. At the lowest quantiles of the distribution, the income-cholesterol association is positive, while from the 75th percentile of cholesterol distribution, the relationship turns to be negative, albeit not statistically significant. More precisely, Figure 1 shows that the "saddle" point is located at the 80th percentile, corresponding to cholesterol equal to 6.5, just a bit higher than the clinical cutpoint of 5. After this threshold, the income gradient increases in magnitude and reaches a peak at the 95th percentile. Interestingly, this pattern indicates that the income gradient rises exactly when cholesterol levels exceed the normal range and are indicative of a disease state.

With respect to the other covariates, we find higher cholesterol levels among the elderly, (especially women) and a high association with education at all levels of cholesterol distribution. However, while the age effect appears to be marginally decreasing along the cholesterol distribution, education effect is marginally increasing: the cholesterol gradient between educated (at any level) and individuals without formal education (reference category) increases along the cholesterol distribution and reach its peak at the 95th percentile of the cholesterol distribution.

The association between income and glycated haemoglobin (GH) is negative at all quantiles of the distribution but it varies highly in magnitude along the distribution. In this case, OLS estimates provides a poor approximation of this association (column 1 of Table 4), while RIF estimates (columns 2-5 of Table 4) show that the income coefficient at 95th percentile of GH distribution is ten times higher than the income coefficient at the 25th percentile (-0.204 vs -0.025). Figure 2 actually shows that income gradient reaches his peak at the 95th percentile of the GH distribution, corresponding to the clinical diagnosis of diabetes (GH >6.5). A similar pattern is observed also for the other covariates: both education and age effects increase at the highest quantiles of GH distribution and they are particularly high around the clinical threshold.

With respect to fibrinogen (Table 5), we find a pattern very similar to the income-GH relationship. We find a negative association at all quantiles of fibrinogen distribution and an higher income gradient at top quantiles. However, the income gradient is smoother because

the income coefficient at the 95th percentile is "only" twice as much as the income coefficient at the 25th percentile (-0.067 vs -0.028). In this case, OLS regression (column 1 of Table 5) provides a good approximation of the average relationship between income and fibrinogen (the OLS coefficient is -0.039). Consistently with the other biomarkers analyzed, we observe that the income gradient is high around the abnormal threshold of fibrinogen, around the 75th percentile (see Figure 3). The income gradient reaches its peak at the 95th percentile of fibrinogen corresponding to fibrinogen values of 4.3, one point more than the upper bound of the normal ranges of fibrinogen. As far as the other variables are concerned, Table 5 shows that both age and education effects increase along the distribution of fibrinogen, in a manner consistent with the other biomarkers analysed.

With respect to ferritin, we find that a positive income gradient at all points of the distribution (Table 6). This is consistent with the fact that higher ferritin values generally indicate better health, with the exception of very extreme values which may also indicate the presence of some health problems. The income-ferritin relationship reflects exactly this pattern. Indeed, we find that income gradient increases up to the 75th percentile and it reduces in magnitude after this threshold, becoming hardly statistically significant at the 95th percentile of ferritin distribution. This pattern is more clearly depicted in Figure 4. The income coefficient increases almost linearly up to the 75th percentile (the income coefficient at the 75th percentile is three times higher than the coefficient at the 25th percentile). After this threshold (corresponding to ferritin level of 100), there is less clear income-ferritin pattern because income coefficient fluctuates widely (but it is always positive) along the ferritin distribution. Also the other covariates exhibit a similar pattern: both age and education effects reduce highly in magnitude at the top quantiles of the ferritin distribution and the education gradient disappears at the 95th percentile of ferritin distribution.

4.2 Oaxaca-Blinder decomposition of gender differentials in biomarkers

The results of Oaxaca-Blinder (OB) decomposition at the 25th, 50th, 75th, 90th and 95th percentile distribution of cholesterol, glycated haemoglobin, fibrinogen and ferritin are shown in Tables 7-10, respectively. Decomposition is expressed always as a difference between levels for males "minus" levels for females. Thus a positive (negative) difference means that a given biomarker value is higher (lower) among males. Tables 7-10 includes total differences, the explained and the unexplained part and their respective standard errors. Detailed decomposition is shown in Figures 5-8, where we highlight the contribution of main factors (income, education, demographics and year fixed effects) to the explained and unexplained part at 25th, 50th, 75th, 90th and 95th percentile distribution of each biomarker.

We illustrate and explain the mechanics of the decomposition using cholesterol results as an example and discuss the main findings for all other biomarkers. Table 7 shows that females have generally higher values of cholesterol at all quantiles of cholesterol distribution. This is not surprising and this is partly explained by the female sex hormone oestrogen which tends to raise HDL cholesterol (good cholesterol). However, according to our results, gender differentials are not the same along the entire distribution. They are negligible and not

statistically significant at the 25th and 50th percentile of cholesterol distribution while they are high and statistically significant at the 90th (0.22 points is the total gender differential) and 95th percentile (around 0.28 points). The second and third rows of Table 7 shows that this is explained both by a difference in the impact of covariates on cholesterol ("unexplained part") and by compositional differences in covariates ("explained part"). At high levels of cholesterol (ie. at the 75th and 90th percentile) compositional differences are predominant, while at extreme levels of cholesterol (at the 95th percentile), the unexplained part is more important to explain gender differentials.

The detailed contribution of income and the other covariates is presented in Figure 5. The decomposition exercise shows a large contribution of demographics (red bar) at the lowest levels of cholesterol. This contribution is mainly due to a difference in the association of health to demographics while compositional differences are less important. The contribution of education is important especially from the 75th percentile and it is also due to differences in the association of health to education across genders. The contribution of income is predominant from the 25th percentile of the cholesterol distribution and it is particularly high at the extreme levels of cholesterol (95th percentile). Also in the case of income, gender differentials in income-health association ("elasticity effect") are much more important than compositional differences ("compositional effect") to explain total differentials. The interpretation of the elasticity effect deserves more attention. The sign of the elasticity effect of any covariates in our OB decomposition comes from the second term in equation (5) and it depends on two factors*: i.* the sign of the coefficients ($\delta_F$and δ_M) *ii.* the differences in coefficients between the regression on the subsample of males and the subsample of females (δ_F- δ_M). When the coefficients are negative, a positive (negative) elasticity effect arises when the coefficient of female regression is larger (smaller) in magnitude than the coefficient of male regression. On the contrary, when the coefficients are positive, a positive (negative) elasticity effect arises when the coefficient in the female regression is smaller (larger) than the coefficient in the male regression. [6] In the case of cholesterol, as shown in the previous section, the income coefficient in the pooled regression is positive at the lowest quantiles of cholesterol and negative from the 80th percentile of cholesterol. Thus, Figure 5 actually shows that there is an important heterogeneity in the association of cholesterol to income which varies significantly along the cholesterol distribution. From the 25th to the 75th percentile of cholesterol distribution, the positive contribution of income means a higher (positive) association of income to cholesterol among males, while at the 90th percentile, the positive contribution we found means that income has a more protective effect (negative effect) among females. However, it is interesting to observe that at very extreme and dangerous levels of cholesterol (at the 95th percentile) we found a negative contribution which indicates a higher protective effect of income on illness among males.

The OB decomposition of glycated haemoglobin (GH) is reported in Table 8. Results indicate higher values of GH among males along the entire distribution of GH and especially at the extreme levels (at the 95th percentile). The higher prevalence of diabetes among males

[6] The elasticity terms in equation (5) is $\bar{X}_F(\delta_M - \delta_F)$. Thus, with negative (positive) coefficients ($\delta_F$and $\delta_M < 0$), a positive elasticity effect arises iff $\delta_F > \delta_M(\delta_M > \delta_F)$ being the sample mean ($\bar{X}_F$) always positive in our case.

is a recent finding of the medical literature and it is mainly imputed to a higher abdominal visceral fat which represent one of the main risk factor for diabetes (see for instance Perreault et al. 2008). Contrarily to cholesterol, gender differentials in GH are largely due a different association of GH to the covariates (row 3 of table 8), while compositional differences are much less important to explain total gender differentials (row 2). Detailed contribution analysis in Figure 6 suggests a marginal role of demographics and a strong contribution of education (mainly due to an elasticity effect) to gender differentials. With respect to income, Figure 6 indicates that its contribution becomes predominant after the 75th percentile of GH distribution. Also in the case of GH, the contribution of income is largely due to gender-related differences in the association of the biomarker to income which vary significantly along the GH distribution. For instance, while at 90th percentile of GH distribution we found a higher protective effect of income on illness among males, we found the opposite (but less important in magnitude) at the 95th percentile of the distribution.

Table 9 indicates general higher levels of fibrinogen among women especially at low levels of fibrinogen which are not indicative of a pathological state. Differences are largely explained by gender differentials in the association of fibrinogen to covariates while compositional differences are much less important. Detailed decomposition presented in Figure 7 shows that gender differentials in fibrinogen are largely explained by income, and by demographics (to a less extent), while education plays a less important role. Figure 7 also shows that the contribution of income is largely explained by an elasticity effect, a pattern observed also for the other biomarkers. Moreover, similarly to the other biomarkers, the elasticity effect is heterogeneous along the distribution of fibrinogen. Up to the 90th percentile of the fibrinogen distribution, the protective effect of income is higher among females while at very extreme levels of fibrinogen income has a much stronger protective effect among males.

Lastly, we report the OB decomposition of gender differentials in ferritin in Table 10. Results indicate the existence of high ferritin levels among males along the entire distribution and especially at the 95th percentile of ferritin distribution. These differences are partly physiologic and associated to blood losses in menstrual women, for instance. However, decomposition analysis also suggests that differences are largely explained by an elasticity effect while compositional differences are less important. Similarly to the other biomarkers, we found that income is the largest contributing factor to gender differentials and this happens along the entire distribution of ferritin. Again, we found that that income contribution is mostly due to an elasticity effect. This indicates an higher effect of income on ferritin among males at all levels. In substantive terms, these results suggest that income has a more protective effect on males for low ferritin problems while we do not detect meaningful gender differentials at highest levels of ferritin, and the income-ferritin association is rarely significant at those levels (as discussed in the paragraph 4.1).

5. Conclusions

The relationship between income and health is probably one of the most explored topics in health economics. A large literature documents the existence of a positive income gradient which is found in several countries, across different age groups and according to several measures of health status. Despite this large interest, less is known about the income-health relationship at different points of the health distribution. Indeed, studies of the influence of economic conditions on health typically measure the effect of the former on the conditional mean of the health status variable through regression analysis. Analysis based solely on the mean while offering useful information, misses potentially important information in other parts of the distribution. For instance, it does not check for non-linearities in the relationship between income and health across the full conditional distribution. Moreover, it does not permit analysis of the role of economic conditions at the tails of the health distribution which are often associated with large welfare losses for individuals and high costs for the health care system.

This paper fills this gap by offering new evidence on the income-health relationship "beyond the mean" of the health distribution. We use a distributional method proposed in the recent literature, the recentered influence function approach (RIF) of Firpo, Fortin and Lemieux (2009), to estimate income gradients across the full distribution for continuous measures of objective health status: blood-based biomarkers. Moreover, we apply Oaxaca-Blinder decompositions at various quantiles of the biomarker distributions to explain gender differentials in biomarkers in England. We use data from the Health Survey for England and we concentrate on four markers highly predictive of the most relevant chronic diseases in Western Countries: total cholesterol, glycated haemoglobin, fibrinogen and ferritin.

A key advantage of using biomarker data is having a measure of health which is free of reporting bias. Indeed, biomarkers are health measures collected during a professional nurse visit and measured on a continuous scale. So, even if our analysis cannot be considered as causal, the absence of reporting bias rules out one important source of endogeneity in the income-health relationship.

Our analysis makes two important contributions to the existing literature. Firstly, analysis beyond the mean allows us to highlight some aspect of the income-health relationship which are overlooked by standard regression methods. In particular, we find that the income-health relationship is non-linear across the health distribution and that the income gradients appear to be higher at the top quantiles of the biomarker distributions, close to the clinical cutpoints that indicate the presence of disease. For instance, we find that the income gradient at the 95th percentile is ten times higher than income gradient at the 25th percentile of glycated haemoglobin, a marker for diabetes. At the same time, the income gradient at the 95th percentile is twice as much as the gradient at the 25th percentile of the distribution of fibrinogen, a marker of many inflammatory diseases, including cardiovascular ones. The income gradient increases almost linearly up to the 75th percentile of ferritin (a marker of anaemia and other important diseases) and it reduces in magnitude after this threshold. In these cases, the analysis at the mean provides a partial view of the income-health relationship. In the case of cholesterol, analysis at the mean leads to misleading conclusions.

For instance, we found that OLS regression suggests a positive association between income and cholesterol. Instead, RIF regression suggests that at lowest quantiles of the distribution, income-cholesterol association is positive, while from the 75th percentile of cholesterol distribution, the relationship turns to be negative. Also in the case of cholesterol, the "saddle" point is very close to the clinical threshold denoting pathologic cholesterol levels.

A second contribution of our paper is the measurement of the gender differentials in biomarkers and the assessment of the contribution of income (and other covariates) to these differentials. Under the assumption of the *invariance of the conditional distribution* that we discussed in Section 3, these effects could also represent the causal effects of gender on health and of gender on income-health relationship, respectively. We find that, besides some physiological reasons, gender differentials in biomarkers are largely explained by a different association of health to covariates across genders. Importantly, detailed decomposition analysis suggests that the heterogeneity in the effect of income on health across genders accounts for a substantial percentage of the total gender differentials in observed health. Moreover, we find that income-health relationship across genders varies significantly along the biomarker distribution and that this depends on the nature of the biomarker considered. At extreme levels of biomarkers, indicating pathological cardiovascular diseases (ie. cholesterol and fibrinogen), we find a higher protective effect of income on illness among males, while we find that this effect is higher among females at lowest quantiles of the distribution (ie. at the 25th and at the 50th percentile of the distribution). On the contrary, at extreme levels of glycated haemoglobin indicating severe diabetes, we find a higher protective effect of income among females. The same pattern is observed for ferritin.

These results might have important policy implications. If we follow the argument of Van Doorslaer and Koolman (2004), the importance of the gender-related differences in the association of income to health would suggest primarily health policy interventions which operate on the health-gradient. At the same time, it seems that fiscal policy interventions are less relevant as compositional differences are not very important to explain the total gender differentials. Moreover, our analysis suggests that health policy interventions should be differently focused across genders and across the distribution of health. For instance, for the purpose of eliminating socio-economic inequalities in health, it seems important to focus on males in poor economic conditions when considering severe cardiovascular diseases. On the contrary, more attention should be paid to females in poor economic conditions when considering severe diabetes or health problems deriving from high-ferritin, such as liver diseases.

Future research might concentrate on the reasons behind the heterogeneity of income-health relationship along the distribution of health status and across genders. Our results indicate that these aspects should be carefully considered when investigating the impact of income, and of other covariate, on health status.

References

American Diabetes Association (ADA), (2001). Executive summary: standards of medical car in diabetes – 2010. Diabetes Care 33 (Suppl 1): S4-S10

Bago d'Uva, T., van Doorslaer, E., Lindeboom, M., O'Donnell, O., (2008), Does reporting heterogeneity bias the measurement of health disparities?, Health Economics, 17, 3: 351-375.

Banks, J., Marmot, M., Oldfield, Z., Smith, J.P., (2006), Disease and Disadvantage in the United States and in England, Journal of American Medical Association, 3, 295, 17: 2037-45.

Bitler, M.P., Gelbach, J.B., Hoynes, H.W., (2006), What mean impacts miss: Distributional effects of welfare reform experiments, The American Economic Review 96: 988-1012.

Blinder, A. S. (1973), Wage Discrimination: Reduced Form and Structural Estimates, The Journal of Human Resources 8: 436–455.

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

Cutler, D., Deaton, A., Lleras-Muney, A., (2006), The Determinants of Mortality, Journal of Economics Perspectives, 20, 3: 97-120.

Dowd, J.B., Goldman, N., (2006), Do biomarkers of stress mediate the relation between socioeconomic status and health?, Journal of Epidemiology and Community Health, 60: 633-639.

Ettner, S.L., (1996), New evidence on the relationship between income and health, Journal of Health Economics, 15,1: 67-85.

Firpo, S., Fortin, N, Lemieux, T., (2009), Unconditional Quantile Regressions, Econometrica, 77, 3: 953-973.

Fortin N., Lemieux T., Firpo S., (2011), Decomposition methods in economics. volume 4, Part A of Handbook of Labor Economics. Elsevier, 1- 102.

Gruenewald TL, Seeman TE, Ryff CD, et al., (2006), Combinations of biomarkers predictive of later life mortality. *PNAS,* 103:14158-14163.

Holland, P.W., (1986), Statistics and Causal Inference, Journal of the American Statistical Association, 81, 396: 945-960.

Jann, B., (2008), The Blinder-Oaxaca decomposition for linear regression models, The Stata Journal, 8 (4): 453-79.

Jones, A., Lomas, J, Rice, N., (2015), Going Beyond the Mean in Healthcare Cost Regressions: a Comparison of Methods for Estimating the Full Conditional Distribution, Health Economics, 24: 1192-1212.

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

Juerges, H., Kruk, E., Reinhold, S., (2013), The effect of compulsory schooling on health -- evidence from biomarkers, Journal of Population Economics, 26: 645-672.

Kakwani, N. C., A. Wagstaff, and E. van Doorslaer, (1997), Socioeconomic Inequalities in Health: Measurement, Computation and Statistical Inference, Journal of Econometrics 77(1): 87–104.

Mackenbach, J.P., Martikainen, P., Looman, C.W., Dalstra, J.A., Lahelma, E., SWdHA working group, (2005), The shape of the relationship between income and self-assessed health: an international study, International Journal of Epidemology, 34, 2: 286-293.

Muennig, P., Sohler, N., Mahato, B., (2007), Socioeconomic status as an independent predictor of physiological biomarkers of cardiovascular disease: Evidence from NHANES, Preventive Medicine, 45: 35-40.

National Institute of Health, (2001), Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 69: 89–95.

Neumark, D., (1988), Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. The Journal of Human Resources 23: 279–295

Oaxaca, R., (1973), Male-Female Wage Differentials in Urban Labor Markets. International Economic Review 14: 693–709.

Perreault, L., Ma, Y., Dagogo-Jack, S., *et al* (2008) Sex Differences in Diabetes Risk and the Effect of Intensive Lifestyle Modification in the Diabetes Prevention Program. Diabetes care 31: 1416-1421.

Ploubidis, G.B., Benova, L., Grundy, E., Laydon, D., DeStavola, B., (2014), Lifelong Socio Economic Position and biomarkers of later life health: Testing the contribution of competing hypotheses, Social Science and Medicine, 119: 258-265.

Hernández-Quevedo, C., Jones, A.M., Rice, N., (2004), Reporting bias and heterogeneity in self-assessed health. Evidence from the British Household Panel Survey, Discussion Papers 18/2004, University of York.

Sattar, N., Murra, H.M., Welsh, P., Blauw, G.J., Buckley, B.M., Cobbe, S., de Craen, A.J.M., Lowe, G.D., Jukema, J.W., Macfarlane, P.W., Murphy, M.B., Stott, D.J., Westendorp, R.G.J., (2009), Are markers of inflammation more strongly associated with risk for fatal than for nonfatal vascular events? Plos Medicine, 6:1-10.

Sen. A., (2002), Health: Perception Vs Observation, British Medical Journal, 13, 324: 860-61.

Sloczynski, T., The Oaxaca–Blinder Unexplained Component as a Treatment Effects Estimator, Oxford Bulletin of Economics and Statistics, 77, 4: 588-604.

Van Doorslaer, E. and Koolman, X., (2004): Exploring the differences in income-related health inequalities across European countries, Health Economics 13, 609–628.

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.

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 Organisation (WHO) (2011a). Use of glycated haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus. Geneva: World Health Organisation

World Health Organisation (WHO) (2011b) Serum ferritin concentrations for the assessment of iron status and iron deficiency in populations. Vitamin and Mineral Nutrition Information System. Geneva: World Health Organization.

Tables and Figures

Table 1. Descriptive Statistics - Biomarkers

Variable	Mean	Std Dev.	Waves[a]	Observations
Cholesterol	5.57	1.15	2,3,4,5,6,7,8,9,10,11	30,770
Glycated haemoglobin	5.58	0.75	2,3,4,5,6,7,8,9,10,11	34,831
Fibrinogen	3.01	0.74	2,3,4,5,8	15,530
Ferritin	99.29	109.3	1,3,4,5,8	13,849

[a] 2002=wave 1; 2012=wave11

Table 2. Descriptive Statistics - Independent Variables

Variable	Mean	Std Dev.
Log (Household Equivalised Income)	10.04	0.82
M (11-18)- Omitted Category	0.01	0.13
M (18-34)	0.09	0.29
M (35-44)	0.10	0.30
M (45-64)	0.16	0.36
M (65-74)	0.06	0.23
M (75+)	0.03	0.17
F (11-18)	0.02	0.13
F (18-34)	0.11	0.31
F (35-44)	0.12	0.33
F (45-64)	0.20	0.40
F (65-74)	0.07	0.25
F (75+)	0.04	0.20
No Qualification- Omitted Category	0.21	0.41
Degree- NVQ 4,5	0.23	0.42
Higher Education	0.12	0.33
NVQ 3- GCE A	0.14	0.35
NVQ 2 – GCE 0	0.23	0.42
NVQ1 – CSE	0.05	0.21
Other Qualifications	0.02	0.15

Table 3. Cholesterol Results - OLS and RIF regressions

	OLS	RIF Regressions				
Variables		***Q25***	***Q50***	***Q75***	***Q90***	***Q95***
Log Income	0.035***	0.048***	0.039***	0.015	-0.014	-0.015
	0.008	*0.013*	*0.011*	*0.013*	*0.017*	*0.023*
M (18-34)	0.695***	1.018***	0.572***	0.388***	0.288***	0.263***
	0.038	*0.087*	*0.044*	*0.039*	*0.040*	*0.054*
M (35-44)	1.353***	1.739***	1.268***	1.058***	0.880***	0.815***
	0.039	*0.085*	*0.045*	*0.046*	*0.057*	*0.074*
M (45-64)	1.559***	1.949***	1.525***	1.276***	1.051***	0.964***
	0.037	*0.082*	*0.041*	*0.041*	*0.050*	*0.069*
M (65-74)	1.395***	1.857***	1.364***	1.081***	0.824***	0.593***
	0.045	*0.086*	*0.052*	*0.057*	*0.077*	*0.102*
M (75+)	1.082***	1.552***	1.013***	0.737***	0.424***	0.248**
	0.050	*0.094*	*0.059*	*0.064*	*0.083*	*0.103*
F (11-18)	0.104**	0.270**	0.022	0.008	0.024	0.067
	0.042	*0.105*	*0.057*	*0.043*	*0.051*	*0.075*
F (18-34)	0.514***	0.842***	0.348***	0.197***	0.159***	0.147***
	0.036	*0.085*	*0.039*	*0.033*	*0.034*	*0.047*
F (35-44)	0.940***	1.451***	0.773***	0.476***	0.392***	0.358***
	0.038	*0.085*	*0.042*	*0.037*	*0.043*	*0.058*
F (45-64)	1.661***	1.988***	1.586***	1.449***	1.240***	1.238***
	0.037	*0.082*	*0.040*	*0.039*	*0.050*	*0.071*
F (65-74)	1.985***	2.098***	1.923***	2.026***	1.848***	1.840***
	0.043	*0.083*	*0.045*	*0.056*	*0.086*	*0.120*
F (75+)	1.787***	2.013***	1.749***	1.646***	1.549***	1.389***
	0.045	*0.087*	*0.052*	*0.066*	*0.102*	*0.140*
Degree- NVQ 4,5	-0.080***	-0.016	-0.087***	-0.090***	-0.186***	-0.225***
	0.021	*0.030*	*0.028*	*0.035*	*0.049*	*0.066*
Higher Education	-0.016	0.016	-0.017	0.016	-0.080	-0.192**
	0.023	*0.031*	*0.031*	*0.040*	*0.058*	*0.076*
NVQ 3- GCE A	-0.094***	-0.088**	-0.059**	-0.091***	-0.154***	-0.223***
	0.021	*0.035*	*0.030*	*0.035*	*0.050*	*0.066*
NVQ 2 – GCE 0	-0.048**	-0.051*	-0.029	0.010	-0.089*	-0.170***
	0.019	*0.028*	*0.026*	*0.033*	*0.048*	*0.064*
NVQ1 – CSE	-0.089***	-0.048	-0.089**	-0.049	-0.223***	-0.317***
	0.030	*0.046*	*0.041*	*0.051*	*0.068*	*0.086*
Other Qualifications	0.009	-0.012	-0.035	-0.013	-0.031	0.058
	0.045	*0.048*	*0.057*	*0.080*	*0.127*	*0.182*
Constant	4.274***	2.851***	4.142***	5.580***	7.178***	7.976***
	0.088	*0.151*	*0.117*	*0.142*	*0.202*	*0.274*

Years Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	30770	30770	30770	30770	30770	30770

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

Table 4. Glycated Haemoglobin Results - OLS and RIF regressions

	OLS	**RIF Regressions**				
Variables		***Q25***	***Q50***	***Q75***	***Q90***	***Q95***
Log Income	-0.065***	-0.025***	-0.037***	-0.048***	-0.094***	-0.204***
	0.005	*0.004*	*0.004*	*0.005*	*0.010*	*0.026*
M (18-34)	0.068***	0.034	0.040*	0.044**	0.117***	0.267***
	0.025	*0.032*	*0.021*	*0.021*	*0.025*	*0.056*
M (35-44)	0.278***	0.221***	0.221***	0.245***	0.289***	0.581***
	0.026	*0.032*	*0.021*	*0.022*	*0.029*	*0.069*
M (45-64)	0.513***	0.338***	0.379***	0.465***	0.649***	1.337***
	0.025	*0.031*	*0.020*	*0.021*	*0.031*	*0.078*
M (65-74)	0.679***	0.381***	0.475***	0.639***	1.116***	2.307***
	0.027	*0.031*	*0.021*	*0.025*	*0.048*	*0.133*
M (75+)	0.697***	0.397***	0.526***	0.742***	1.187***	2.137***
	0.030	*0.031*	*0.022*	*0.027*	*0.059*	*0.166*
F (11-18)	-0.032	-0.027	-0.009	-0.041*	-0.064***	-0.134***
	0.028	*0.036*	*0.023*	*0.021*	*0.017*	*0.036*
F (18-34)	0.034	-0.019	-0.004	0.022	0.106***	0.262***
	0.024	*0.031*	*0.020*	*0.020*	*0.022*	*0.051*
F (35-44)	0.158***	0.109***	0.099***	0.116***	0.177***	0.449***
	0.025	*0.032*	*0.020*	*0.021*	*0.025*	*0.060*
F (45-64)	0.411***	0.310***	0.333***	0.396***	0.475***	0.839***
	0.024	*0.031*	*0.020*	*0.021*	*0.027*	*0.065*
F (65-74)	0.610***	0.398***	0.497***	0.661***	0.915***	1.638***
	0.027	*0.031*	*0.021*	*0.024*	*0.044*	*0.118*
F (75+)	0.628***	0.407***	0.537***	0.738***	1.100***	1.618***
	0.028	*0.032*	*0.022*	*0.026*	*0.053*	*0.137*
Degree- NVQ 4,5	-0.154***	-0.079***	-0.083***	-0.133***	-0.243***	-0.558***
	0.013	*0.010*	*0.009*	*0.012*	*0.026*	*0.073*
Higher Education	-0.119***	-0.038***	-0.047***	-0.089***	-0.186***	-0.452***
	0.014	*0.010*	*0.009*	*0.013*	*0.030*	*0.083*
NVQ 3- GCE A	-0.154***	-0.071***	-0.077***	-0.143***	-0.238***	-0.553***
	0.013	*0.011*	*0.010*	*0.012*	*0.026*	*0.073*
NVQ 2 – GCE 0	-0.107***	-0.042***	-0.044***	-0.088***	-0.195***	-0.480***
	0.012	*0.009*	*0.008*	*0.011*	*0.025*	*0.070*
NVQ1 – CSE	-0.094***	-0.018	-0.023*	-0.071***	-0.167***	-0.365***
	0.019	*0.014*	*0.013*	*0.018*	*0.040*	*0.120*
Other Qualifications	-0.116***	-0.018	-0.045***	-0.092***	-0.257***	-0.533***

	0.027	*0.018*	*0.017*	*0.024*	*0.058*	*0.162*
Constant	6.043***	5.407***	5.657***	6.126***	6.827***	8.110***
	0.055	*0.052*	*0.042*	*0.051*	*0.101*	*0.277*
Years Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	34831	34831	34831	34831	34831	34831

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

Table 5. Fibrinogen Results - OLS and RIF regressions

	OLS	**RIF Regressions**				
Variables		***Q25***	***Q50***	***Q75***	***Q90***	***Q95***
Log Income	-0.039***	-0.028***	-0.050***	-0.057***	-0.041***	-0.067***
	0.008	*0.010*	*0.011*	*0.014*	*0.016*	*0.025*
M (18-34)	0.053	0.026	-0.010	0.035	0.119*	0.113
	0.039	*0.070*	*0.058*	*0.060*	*0.070*	*0.102*
M (35-44)	0.247***	0.266***	0.202***	0.163***	0.183***	0.168*
	0.040	*0.069*	*0.058*	*0.060*	*0.070*	*0.101*
M (45-64)	0.447***	0.470***	0.438***	0.353***	0.358***	0.392***
	0.039	*0.067*	*0.057*	*0.060*	*0.072*	*0.106*
M (65-74)	0.635***	0.586***	0.625***	0.589***	0.622***	0.618***
	0.045	*0.069*	*0.062*	*0.074*	*0.093*	*0.147*
M (75+)	0.839***	0.672***	0.795***	0.846***	0.887***	1.316***
	0.049	*0.069*	*0.065*	*0.084*	*0.117*	*0.208*
F (11-18)	0.143***	0.144**	0.103	0.046	0.148	0.107
	0.043	*0.073*	*0.074*	*0.080*	*0.108*	*0.150*
F (18-34)	0.385***	0.428***	0.359***	0.329***	0.358***	0.316***
	0.038	*0.066*	*0.057*	*0.060*	*0.073*	*0.104*
F (35-44)	0.402***	0.407***	0.406***	0.385***	0.346***	0.312***
	0.039	*0.068*	*0.058*	*0.062*	*0.072*	*0.103*
F (45-64)	0.581***	0.571***	0.584***	0.552***	0.488***	0.482***
	0.039	*0.067*	*0.056*	*0.061*	*0.073*	*0.106*
F (65-74)	0.798***	0.701***	0.784***	0.852***	0.772***	0.861***
	0.044	*0.067*	*0.061*	*0.073*	*0.095*	*0.150*
F (75+)	0.880***	0.689***	0.882***	1.026***	0.944***	1.120***
	0.046	*0.068*	*0.062*	*0.082*	*0.113*	*0.178*
Degree- NVQ 4,5	-0.193***	-0.157***	-0.193***	-0.262***	-0.279***	-0.222***
	0.020	*0.024*	*0.026*	*0.035*	*0.042*	*0.067*
Higher Education	-0.127***	-0.082***	-0.119***	-0.178***	-0.223***	-0.194**
	0.021	*0.025*	*0.029*	*0.039*	*0.048*	*0.077*
NVQ 3- GCE A	-0.127***	-0.088***	-0.128***	-0.187***	-0.239***	-0.215***
	0.021	*0.026*	*0.029*	*0.038*	*0.044*	*0.067*
NVQ 2 – GCE 0	-0.096***	-0.076***	-0.112***	-0.148***	-0.162***	-0.146**

	0.018	*0.021*	*0.024*	*0.034*	*0.043*	*0.066*
NVQ1 – CSE	-0.032	-0.039	-0.032	-0.038	-0.103	-0.056
	0.028	*0.033*	*0.037*	*0.051*	*0.065*	*0.104*
Other Qualifications						
	-0.118***	-0.042	-0.098**	-0.194***	-0.195**	-0.203
	0.039	*0.036*	*0.050*	*0.070*	*0.092*	*0.143*
Constant	3.135***	2.640***	3.190***	3.738***	3.954***	4.448***
	0.083	*0.118*	*0.119*	*0.149*	*0.169*	*0.259*
Years Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	15530	15530	15530	15530	15530	15530

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

Table 6. Ferritin Results- OLS and RIF regressions

	OLS	**RIF Regressions**				
Variables		***Q25***	***Q50***	***Q75***	***Q90***	***Q95***
Log Income	2.988***	0.530***	0.795***	1.301***	0.994***	1.412*
	1.077	*0.093*	*0.150*	*0.295*	*0.373*	*0.734*
M (18-34)	65.895***	4.381***	14.167***	18.342***	9.725***	10.611***
	4.093	*0.388*	*0.624*	*1.077*	*1.067*	*1.862*
M (35-44)	93.800***	4.272***	15.325***	23.245***	19.288***	29.210***
	4.647	*0.430*	*0.669*	*1.290*	*1.611*	*3.276*
M (45-64)	99.457***	4.003***	14.269***	23.593***	21.836***	29.946***
	4.270	*0.439*	*0.660*	*1.117*	*1.304*	*2.399*
M (65-74)	100.438***	3.900***	14.553***	25.315***	24.274***	33.142***
	4.887	*0.468*	*0.718*	*1.303*	*1.718*	*3.570*
M (75+)	73.658***	2.832***	11.505***	18.698***	16.438***	25.636***
	5.432	*0.497*	*0.796*	*1.523*	*1.744*	*3.407*
F (11-18)	-11.592***	-3.344***	-2.828***	-1.098***	0.019	0.071
	3.776	*0.429*	*0.425*	*0.410*	*0.199*	*0.183*
F (18-34)	-10.315***	-2.451***	-0.939	-2.467***	-1.179**	0.404
	3.924	*0.428*	*0.574*	*0.739*	*0.591*	*0.969*
F (35-44)	-5.978	-2.110***	-0.213	-1.954**	-0.580	0.861
	4.537	*0.495*	*0.685*	*0.875*	*0.715*	*1.161*
F (45-64)	22.754***	1.110**	6.235***	5.023***	2.630***	3.753***
	4.218	*0.453*	*0.670*	*0.912*	*0.728*	*1.119*
F (65-74)	48.474***	2.727***	10.312***	12.039***	8.250***	11.069***
	4.879	*0.458*	*0.711*	*1.180*	*1.226*	*2.207*
F (75+)	33.014***	2.043***	8.676***	8.388***	4.537***	6.464***
	4.973	*0.495*	*0.783*	*1.300*	*1.173*	*2.018*
Degree- NVQ 4,5	5.888**	0.593***	0.648*	1.499*	1.996**	0.567
	2.820	*0.221*	*0.376*	*0.770*	*0.987*	*1.849*

Higher Education	7.426**	0.303	0.995**	1.558*	1.756	1.843
	3.110	*0.237*	*0.412*	*0.869*	*1.159*	*2.262*
NVQ 3- GCE A	-2.454	0.330	0.565	-0.377	-1.508*	-2.744
	2.934	*0.239*	*0.402*	*0.788*	*0.905*	*1.688*
NVQ 2 – GCE 0	-1.808	0.908***	0.714**	-0.733	-1.783**	-2.417*
	2.431	*0.199*	*0.337*	*0.623*	*0.726*	*1.355*
NVQ1 – CSE	-2.170	0.493	0.762	0.661	-0.440	0.180
	4.143	*0.319*	*0.530*	*1.135*	*1.362*	*2.471*
Other Qualifications	1.044	0.397	0.905	1.581	0.246	-4.641*
	6.639	*0.509*	*0.893*	*1.681*	*1.759*	*2.412*
Constant	15.263	26.263***	46.697***	92.894***	176.526***	239.064***
	10.748	*0.940*	*1.500*	*2.869*	*3.614*	*7.095*
Years Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	14188	14188	14188	14188	14188	14188

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

Table 7. Oaxaca-Blinder Decomposition of gender differentials in Cholesterol

	Q25	**%**	**Q50**	**%**	**Q75**	**%**	**Q90**	**%**	**Q95**	**%**
Δ_{M-F}	-0.010		0.008		-0.068**		-0.225***		-0.276***	
	0.019		*0.017*		*0.021*		*0.026*		*0.032*	
Compositional	-0.028	281	-0.047***	-555	-0.055***	-647	-0.056***	81	-0.056***	20
	0.006		*0.008*		*0.009*		*0.009*		*0.001*	
Elasticity	0.018	-180	0.055**	655	-0.013	-153	-0.168***	19	-0.219***	80
	0.018		*0.165*		*0.019*		*0.025*		*0.031*	

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

Table 8. Oaxaca-Blinder Decomposition of gender differentials in Glycated Haemoglobin

	Q25	**%**	**Q50**	**%**	**Q75**	**%**	**Q90**	**%**	**Q95**	**%**
Δ_{M-F}	0.033***		0.032***		0.024**		0.073**		0.336***	
	0.006		*0.005*		*0.007*		*0.014*		*0.041*	
Compositional	-0.009***	-27	-0.015***	-47	-0.022***	-91	-0.030***	-41	-0.060***	-18
	0.002		*0.003*		*0.004*		*0.005*		*0.009*	
Elasticity	0.042***	127	0.047***	147	0.046***	191	0.103***	141	0.397***	118
	0.006		*0.005*		*0.007*		*0.013*		*0.040*	

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

Table 9. Oaxaca-Blinder Decomposition of gender differentials in Fibrinogen

	Q25	%	Q50	%	Q75	%	Q90	%	Q95	%
Δ_{M-F}	-0.330***		-0.25***		-0.381***		-0.244***		-0.103**	
	0.014		*0.014*		*0.019*		*0.029*		*0.04*	
Compositional	-0.022***	6	-0.027***	11	-0.033***	9	-0.034***	14	-0.033***	32
	0.004		*0.004*		*0.006*		*0.007*		*0.009*	
Elasticity	-0.307***	93	-0.223***	89	-0.348***	91	-0.209***	86	-0.069*	68
	0.014		*0.014*		*0.018*		*0.028*		*0.039*	

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

Table 10. Oaxaca-Blinder Decomposition of gender differentials in Ferritin

	Q25	%	Q50	%	Q75	%	Q90	%	Q95	%
Δ_{M-F}	28.97***		53.99***		84.13***		125.04***		159.86***	
	0.14		*0.217*		*0.414*		*0.917*		*1.399*	
Compositional	0.004	0.1	-0.046	-0.01	-0.152	-0.01	-0.248	-0.2	-0.282	-0.2
	0.028		*0.047*		*0.110*		*0.161*		*0.246*	
Elasticity	28.96***	99.9	54.04***	100.09	84.29***	100.09	125.29***	100.2	160.14***	100.2
	0.14		*0.210*		*0.402*		*0.905*		*1.389*	

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

Figure 1. Income Coefficient of RIF regression. Cholesterol

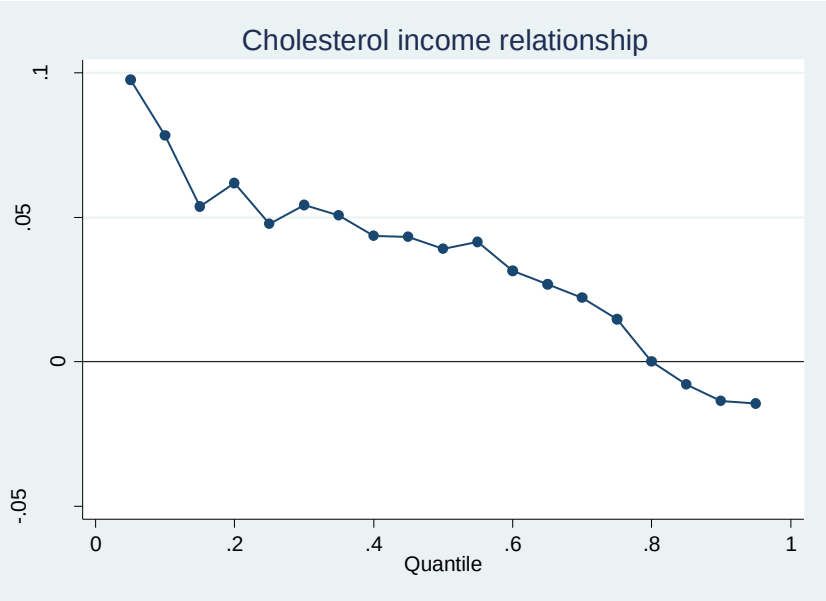

Figure 2. Income Coefficient of RIF regression. Glycated Haemoglobin

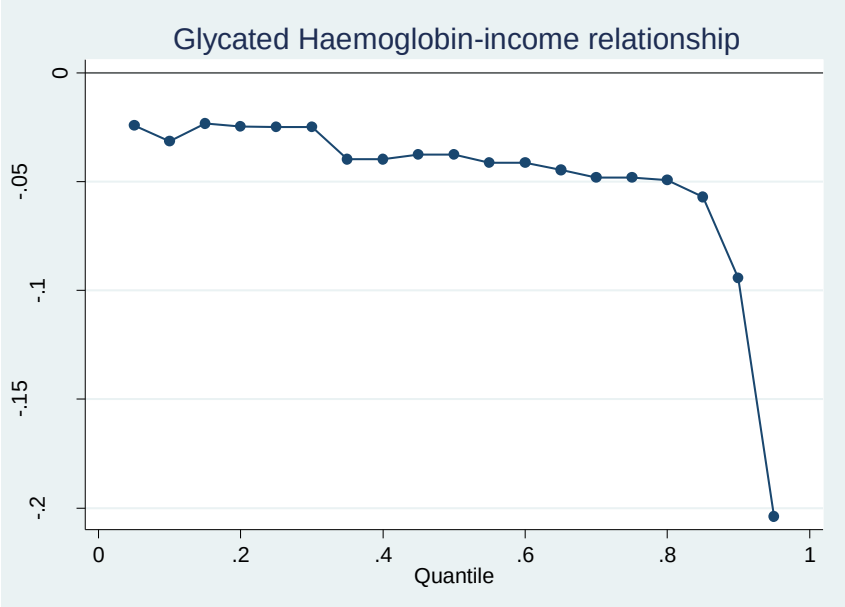

Figure 3. Income Coefficient of RIF regression. Fibrinogen

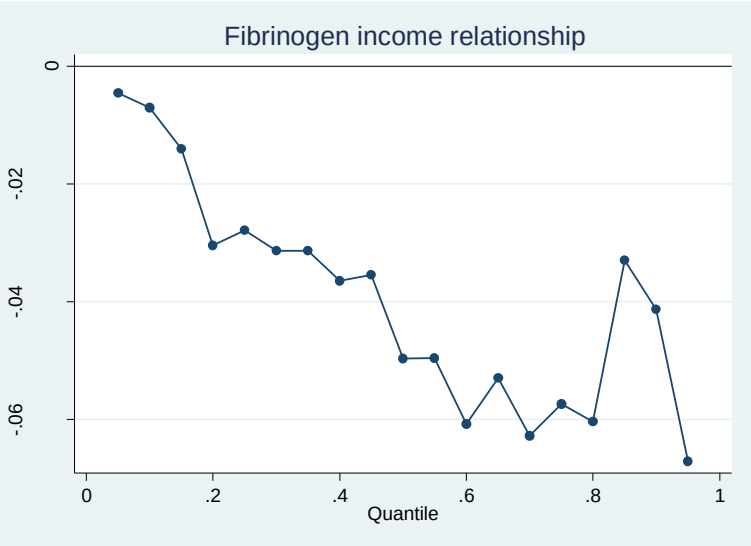

Figure 4. Income Coefficient of RIF regression. Ferritin

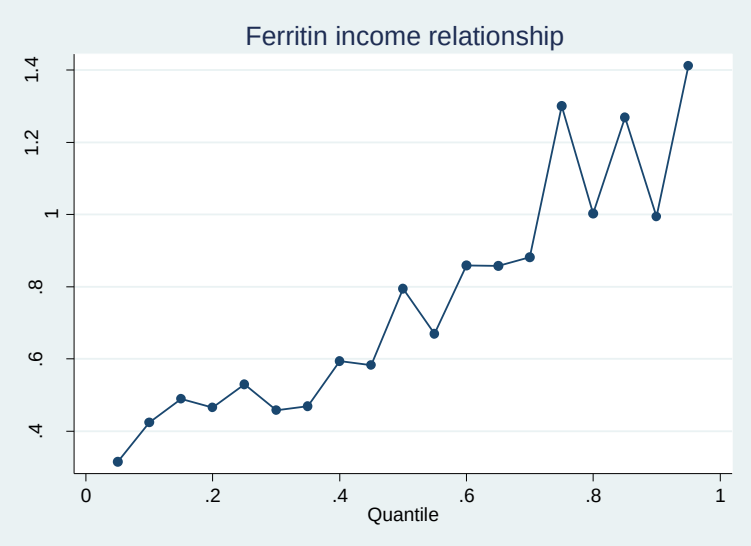

Figure 5. Detailed Decomposition of gender differentials- Cholesterol

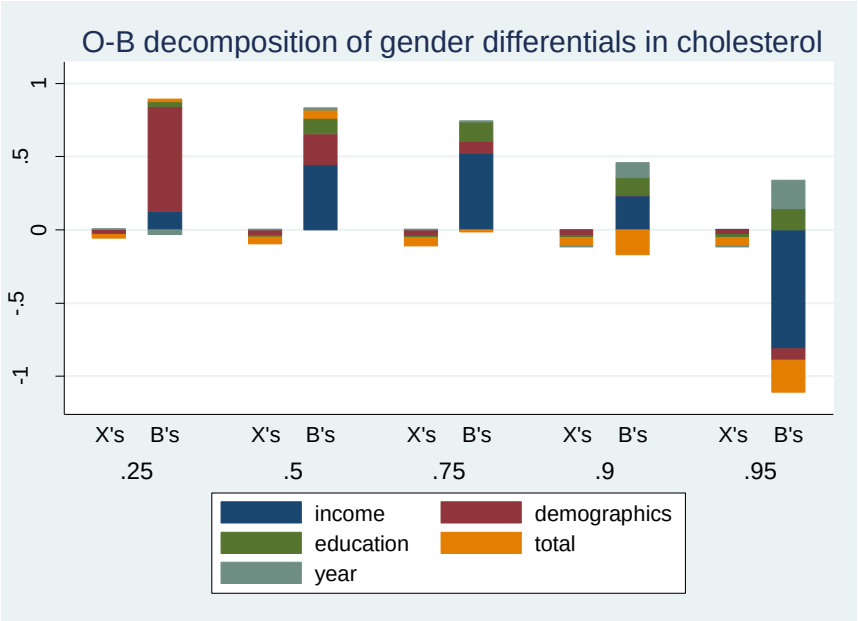

Figure 6. Detailed Decomposition of gender differentials- Glycated Haemoglobin

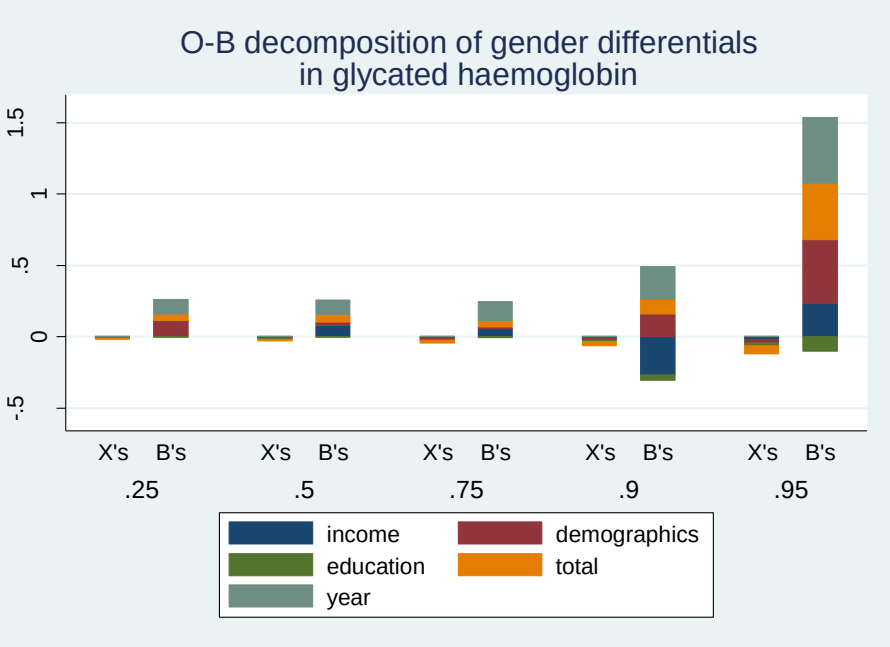

Figure 7. Detailed Decomposition of gender differentials- Fibrinogen

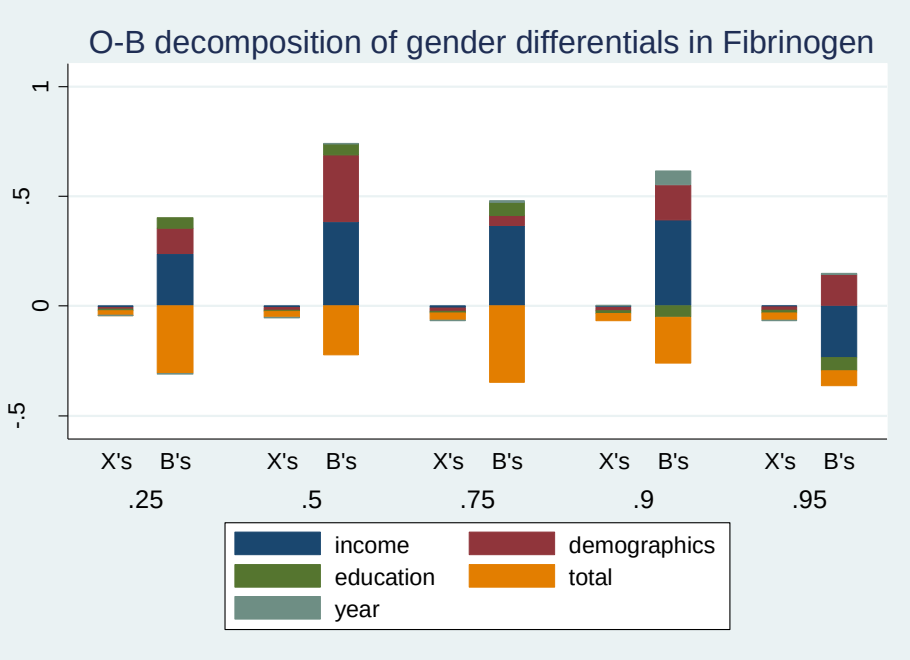

Figure 8. Detailed Decomposition of gender differentials- Ferritin