LUND UNIVERSITY

Do Socioeconomic Factors Really Explain Income-Related Inequalities in Health? Applying a Twin Design to Standard Decomposition Analysis

Gerdtham, Ulf; Lundborg, Petter; Nystedt, Paul; Lyttkens, Carl Hampus

2012

Do Socioeconomic Factors Really Explain Income-Related Inequalities in Health?
Applying a Twin Design to Standard Decomposition Analysis

Ulf-G Gerdtham
Petter Lundborg
Carl Hampus Lyttkens
Paul Nystedt

August 2012

LUND
UNIVERSITY

Do Socioeconomic Factors Really Explain Income-Related Inequalities in Health? Applying a Twin Design to Standard Decomposition Analysis

Gerdtham, U-G*[a,b,c], Lundborg, P.[a,b,d,e], Lyttkens, C.H. [a,b], Nystedt, P.[b,f]

[a]Department of Economics, Lund University

[b]Health Economics Program, Lund University

[c]Centre for Primary Health Care Research, Lund University

[d]Tinbergen Institute

[e]IZA Bonn

[f]Department of Economics, University of Linköping

Abstract

The concentration index and decomposition analysis are commonly used in economics to measure and explain socioeconomic inequalities in health. Such analysis builds on the strong assumption that a health production function can be estimated without substantial bias implying that health is caused by socioeconomic outcomes, which is hard to prove. This article contributes to the decomposition literature by applying a twin design to standard decomposition analysis of socioeconomic health inequalities in Sweden. The twin-based decomposition estimates, which control for unobserved endowments at the twin-pair level, are much lower in magnitude than estimates obtained via typical OLS on the same sample. This demonstrates that OLS-based decompositions are severely upward biased due to underlying confounders, exaggerating the contribution of income and education to health inequality, which in turn limits the usefulness of such decompositions for policy purposes.

* Correspondence: Ulf-G Gerdtham,
Department of Economics,
P. O. Box 7082,
SE-220 07 Lund, Sweden,
phone: +46 46 2224810, fax: +46 46 2224118,
email: ulf.gerdtham@nek.lu.se

1. Introduction

Numerous studies report a strong socioeconomic gradient in health and longevity, regardless of the population studied and regardless of how socioeconomic status and health are measured (see Ettner, 1996; Bloom and Cunning, 2000; Smith, 1999; Benzeval and Judge, 2001; Deaton, 2003; Gerdtham and Johannesson, 2000, 2002, 2004; Baum and Ruhm, 2009). Despite improvements in average health status over recent decades, this health gradient has persisted and even increased in most western countries (Mackenbach et al., 2003; van Doorslaer and Koolman, 2004; Kunst et al., 2005; Shkolnikov et al., 2011). Thus, reducing the magnitude of health inequality still poses a challenge, and has become a major policy objective for many European governments (Marmot et al., 2010).

Recently, the Marmot Review suggested that the strong association between health and socioeconomic status implies that socioeconomic inequalities have to be reduced in order to reduce socioeconomic inequalities in health (Marmot et al., 2010; Marmot, 2012). Accordingly, policy measures affecting socioeconomic inequalities, such as income redistribution and publicly financed education, are taken to influence health inequalities. This conclusion is based on the strong assumption that the association between socioeconomic status and health reflects a causal effect running from the former to the latter, which has attracted extensive debate and disagreement in the literature (e.g. Smith 1999; Deaton 2002; Cutler et al., 2008). This lack of consensus reflects the incomplete nature of our knowledge about underlying causal mechanisms and channels behind the disparities in health. Policies aimed at reducing disparities may therefore easily be ineffective, inconsistent, and even counterproductive (Deaton, 2011), and so there is an urgent need to improve our understanding of the origins of socioeconomic health inequalities in order to enable the design of efficient policy.

There is a large and growing body of research in economics on inequalities in health (Wagstaff and Van Doorslaer, 2000; Van Doorslaer and Koolman, 2004; Fleurbaey and Schokkaert, 2011). One of the dominant metrics for measuring socioeconomic inequalities in health is the concentration index, which measures health inequalities in relation to an individual's socioeconomic rank, for example in terms of income (Fleurbaey and Schokkaert, 2011). An attractive feature of this index is its illustrative and intuitive interpretation. The index also takes the whole distribution into account rather than only calculating differences between extremes of the population such as the rich and the poor (Wagstaff and Van Doorslaer, 2000; Kjellsson and Gerdtham, 2011).

In order to provide information on the relative importance of different factors and the magnitude of their individual contributions, Wagstaff et al. (2003) proposed a decomposition method. Here, the concentration index of health is expressed as a function of the means and inequalities of the factors in a health production function as well as the regression effects (elasticities) of health factors (see below). The results of such an analysis could be useful to policy makers in their search for alternative policies to achieve maximum reductions in health inequalities. For example, if lack of education stands out as a prominent source of health inequalities, possibly due to the diffusion of health-related information and knowledge in society, one could target educational programs towards individuals who are disadvantaged with respect to education.

In a widely cited paper, Van Doorslaer and Koolman (2004) employed this method to analyze income-related inequalities in self-rated health in 13 European Union member states based on cross-sectional data. They found income-related health inequalities in all countries; these were particularly salient in Portugal, the UK, and Denmark, and relatively low in the Netherlands and Germany. The authors concluded that income is a major contributor to the health concentration index, along with educational attainment and employment status. This

indicates that the reduction of socioeconomic inequalities, particularly in income, may be a major vehicle to diminish health inequalities. Subsequent research has used this method to analyze the sources of health disparities using data from different countries and various health measures, income definitions, and determinants (e.g. Leu and Schellhorn, 2004; Gomez and Nicholas, 2005; Lauridsen, 2007; Hosseinpoor et al., 2006; Costa-Font et al., 2009; McGrail et al., 2009; Morasae et al., 2012).

For policy purposes, the usefulness of any decomposition analysis critically hinges upon whether the estimated health function reflects causal effects of socioeconomic factors on health. In practice, this is tremendously difficult to prove (Deaton, 2011). There are two fundamental conceptual problems. First, causality may run from inherent health status to, for example, income, rather than the other way around. Second, there may be some third, unobserved factor, perhaps of genetic or environmental origin, that determines both health and income (Fuchs, 1982; Smith, 1999; Mackenbach, 2005; Deaton, 2011). Similar causality issues also hold for education and other socioeconomic factors. One limitation of the majority of prior decomposition studies is that they consider neither reverse causality nor the unobserved endowment and heritability that are correlated with socioeconomic factors. Thus the estimated contributions of key factors may well be confounded with the effects of innate ability and other hard-to-define or hard-to-measure factors relating to personality and family background. Consequently, the estimated coefficients show that health is associated with a number of "factors" which in turn are correlated with income positions, rather than providing any causal inference.

There are few studies that explicitly deal with the endogeneity problem in the context of decomposition analysis of health inequalities (Wildman, 2003; Jones and Lopez Nicolas, 2006; Islam, 2010). In a different context, some studies have tried to isolate the causal effect of one particular socioeconomic factor on health. Ettner (1996), for instance, used state

unemployment rate, work experience, parental education, and spouse characteristics as instruments for family income. Meer et al. (2003) used inheritances and gifts, and Lindahl (2005) monetary lottery prizes to capture exogenous variation in wealth. Other studies have used natural experiments to capture exogenous variation like the reunification of Germany (Frijters et al., 2005) or compulsory education reform in Sweden (Meghir et al., 2012). Some studies have also analyzed changes in health by controlling for past health (Buckley et al., 2004; Gerdtham and Johannesson, 2002) or controlling for panel data fixed effects (e.g. Wildman, 2003; Jones and Schurer, 2007; Islam et al., 2010).

The present article adds to the existing decomposition literature in that it revisits the analysis of the effect of socioeconomic factors on income-related health inequalities, using the twin data design as an identification strategy. The twin approach has been exploited previously in the economics literature, mainly in order to estimate the returns to education (Ashenfelter and Kreuger, 1994; Ashenfelter and Rouse, 1998; Bonjour et al., 2003; Isacsson, 2004). Recently, a number of papers have also used the design to estimate the causal effect of education on health (Lundborg, 2012; Amin and Behrman, 2009; Webbink, 2010; Behrman et al., 2011).

There are two kinds of twins: dizygotic (DZ) and monozygotic (MZ). Except for being born at the same time, DZ twins are genetically just like ordinary siblings in the sense that they are the product of two different eggs. MZ twins are generally denoted "identical" as they emerge from a single egg sharing the same set of genes. Because twins raised together share childhood family environment as well as genetic factors – with 50% of their DNA in common for DZ and 100% for MZ twins – they provide an opportunity to control for unobserved genetic and family background endowment factors. By examining within-twin-pair differences in health due to within-twin-pair differences in socioeconomic factors, the influence of confounding "ability" factors may be removed; that is, factors common to both

twins in a given twin pair, such as higher ability and favorable family background, are differenced out. This may be particularly suitable in this context, since both genes and family background are likely to be important determinants of many socioeconomic factors as well as health.

A twin design brings certain advantages to the analysis of the relation between socioeconomic factors and adult health. First, it allows us to also estimate the influence of factors that do not change over time, such as educational status, by relying on differences in socioeconomic factors between twins at a given point in time. Note that this is in contrast to the fixed effects panel data approach, where identification requires that the variable of interest varies over time; this may be true for certain factors like income, but not for others, such as education. Hence, it is difficult to assess the relative contribution of factors such as education using a fixed effects panel data approach.[1]

Second, twin estimates are likely to reflect the average effect in the population of interest, since differences within twin pairs in factors such as income, which are used to identify the income effect, are likely to be evenly spread across the income distribution. An instrumental variables approach, in contrast, estimates local average treatment effects, where the estimated effect is only identified for the marginal group affected by the instrument. Since the instruments used typically only affect a small subgroup in the population, it is normally not the case that this effect can be generalized to the greater population. Moreover, most instrumental variable studies are only able to instrument for one explanatory variable, such as education, which means that the endogeneity problem remains for the other variables

---

[1] The twin approach, on the other hand, requires that the variable of interest varies within twin pairs. As we will show, this requirement is in general fulfilled for most variables in our analyses.

indicating socioeconomic status. This, again, naturally complicates the assessment of the relative contribution of the different variables indicating socioeconomic status.

The present article uses a unique data set, including data on a vast majority of Swedish native twins born between 1896 and 1958. The data set links survey data on self-reported health, health-related behavior, and labor market information to register data on income and education. In the decomposition analysis, we report results based both on a health production function estimated by OLS, with the health of twins treated as independent observations, and by within-twin-pair (WTP) estimations, which removes the influence of factors shared by twins.

To preview our findings: the OLS-based decomposition indicates, in line with most prior cross-sectional studies, that factors such as income and education explain a significant part of the measured income-related inequalities in health. The WTP-based decomposition, however, indicates much weaker and statistically insignificant contributions of income and education. A reasonable interpretation is that OLS-based decompositions are subject to severe upward biases due to the influence of unobserved factors. In other words, our results suggest that most of the socioeconomic disparities in health are attributed to hard-to-measure factors that affect both socioeconomic status and health, such as genes and early life conditions. From a policy perspective, reducing inequalities in income and education would therefore do much less to reduce socioeconomic disparities in health than what might be expected from OLS-based analyses. In fact, since the WTP estimates can be viewed as an upper bound of the effect of socioeconomic factors on health, our results suggest that factors such as income and education play a very small role in causally explaining socioeconomic disparities in health.

The paper unfolds as follows. Section 2 provides a background discussion and reviews the literature on decomposition analysis of the health concentration index and the value of WTP estimation. Section 3 presents the empirical model and the data material, and explains

the details of the variable constructions. Section 4 reports the results including some comparisons with another data set based on a general population, and section 5 concludes.

2. Methodological considerations

Income-related inequalities in health can be defined as variations in health across individuals with varying income. One popular measure of such (bivariate) inequalities is the concentration index (henceforth *C*), which has as one of its properties that it can be decomposed into various factors of health. Techniques to measure and decompose *C* are summarized below.

Measuring income-related inequalities in health

The measurement of income-related inequalities in health using *C* is discussed in Kakwani *et al.* (1997). In brief, the sample is ranked by income, typically beginning with the least advantaged. Given a continuous measure of health status, a health concentration curve can be constructed which plots the cumulative proportion of the population ranked by the health variable against the cumulative proportion ranked by income. *C* is defined as twice the area between the concentration curve and the diagonal, which represents perfect equality of health across income. A convenient way to estimate *C* for health is:

$$C = \frac{2}{\bar{h}} \text{cov}(h_i, r_i) \qquad (1)$$

where, h_i is the *i*th individual's health, $\bar{h}$ is the mean of health, r_i is the relative fractional income rank of the *i*th individual, and *cov* is the covariance. Thus one can estimate *C* of h by computing twice the covariance between h and the relative rank normalized by the mean of h. If the h variable is continuous, then *C* takes values between −1 and +1 depending on whether

8

the total h is concentrated amongst the least (-1) or most (+1) economically advantaged individuals.

Decomposing the concentration index of health

Wagstaff *et al.* (2003) demonstrate that *C* of a continuous health measure can be decomposed into the contributions of individual health factors, given that it is possible to specify a health function in an additive linear form as:

$$h_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki} + \varepsilon_i \qquad (3)$$

where x_k are health factors and ε is a disturbance term. By using equations (1) and (3), *C* of health can be written as:

$$C_h = \sum_{k=1}^{K} (\beta_k \frac{\overline{x}}{\overline{h}}) \cdot C_k + \frac{2}{\overline{h}} cov(\varepsilon_i, r_i) \qquad (4)$$

The first term in brackets represents the elasticity of h with respect to x_k evaluated at the sample means (x_k and h), and C_k denotes *C* of x_k against income. Thus *C* of h can be decomposed into an "explained part" and an "unexplained part". The "explained" part can be broken down into the contributions of each of the health factors in the production function. The "unexplained" part is a scaled measure of the covariance of the residuals in the regression model with respect to the position of the individual in the income distribution. As such, the unexplained part should be small if the h function contains income as an explanatory variable (Gravelle and Sutton, 2003). Equation (4) highlights the two conditions that must hold for a health factor to make a contribution to inequalities in health. First, it must have a significant effect on health so that changes in the factor in question produce changes in health. Second, the health factor must be distributed unequally across income groups. If both these conditions hold, health inequalities may be subject to policies either targeting the health factor directly or addressing its effects.

Bootstrapping techniques may be used in order to derive standard errors for many of the statistics calculated in the analysis (i.e. the contributions of separate health factors). In this article, the bootstrap estimates for standard errors are computed following Van Doorslaer and Koolman (2002). The number of replications has been set to 1000.

Twin data approach to the regression of health

To see how the twin design may help to reduce the problem of unobserved heterogeneity, consider an individual i, whose health, h_i, is determined by:

$$h_i = \beta y_i + \alpha A_i + u_i, \tag{5}$$

where y_i denotes income (or education or any other socioeconomic factor) and A_i denotes unobserved factors affecting health, such as genetic traits and personal characteristics as well as family background. Next, let income be:

$$y_i = \delta A_i + \xi_i, \tag{6}$$

where ξ_i denotes a income-specific random term. This gives rise to the standard result that an OLS estimate of β is biased such that:

$$\text{plim}(\beta_{OLS}) = \beta + \alpha((\sigma_{Ay})/(\sigma_y^2)). \tag{7}$$

Since the unobserved factors (A_i) are likely to be positively correlated with both income and health, it is usually assumed that an estimate of β_{OLS} will be upward biased. Since the contributions of income and education to *C* in the decomposition analysis are calculated as the (health) elasticity of income and education multiplied by the inequalities of income and education, their contributions will also be upward biased.

In the twin-differencing strategy, let h_{1j} and h_{2j} denote the health of the first and second twin in the jth twin pair. Assume now that the unobserved component is made up of two parts. The first part, μ_j, denotes unobserved factors that vary between MZ twin pairs but

not within pairs, such as genetic characteristics and certain early life environmental factors. Finally, ε_{1j} and ε_{2j} denote unobserved factors specific to each twin. This can be written as:

$$h_{1j} = \beta y_{1j} + \mu_j + \varepsilon_{1j}, \qquad (8)$$

$$h_{2j} = \beta y_{2j} + \mu_j + \varepsilon_{2j}, \qquad (9)$$

Next, we take the difference between (8) and (9):

$$h_{1j} - h_{2j} = \beta_{WTP} (y_{1j} - y_{2j}) + \varepsilon_{1j} - \varepsilon_{2j} \qquad (10)$$

where β_{WTP} is the within-twin-pair, or fixed effects, estimate of the effect of income. Insofar as μ_j captures the influence of common family background, which may be of a genetic or environmental character, their influence will vanish, since they will be differenced out of the equation. This means that an OLS estimate of (10) will no longer be biased due to unobserved twin-pair specific variables.

There are two well-known potential problems with a twin design. First, WTP estimates may still be biased if there are important unobserved differences between the twins that relate to both income and health. As noted by Bound and Solon (1999), even MZ twins may differ in factors such as birth weight, for instance, which has been linked to both adult earnings and education (e.g. Behrman and Rosenzweig, 2004; Black et al., 2007; Royer, 2009).[2] Moreover, there may exist ability differences between MZ twins, so that the twin with greater non-genetically induced ability also obtains better health, irrespective of education or

---

[2] Other studies using the twin design find no significant impact, however (Bonjour et al., 2003; Miller et al., 2005; Petersen et al., 2006). Also note that in most studies that find an effect on schooling, the effects are small in magnitude. In Royer (2009), for instance, an increase in birth weight by 250 grams, which would be quite a policy achievement, only leads to 0.03-0.04 of a year of additional schooling.

income.[3] Such results were obtained by Sandewall et al. (2009), who showed that the difference between MZ twins in cognitive test scores at age 18 was a significant predictor of their later schooling. If cognitive ability has positive effects on both later income and health, our twin-based estimates still risk being biased upward, as we do not have data on cognitive test scores. The results of Sandewall et al. (2009) seem compelling at first sight, but note that the cognitive test scores they rely on are measured at an age where they are likely to have been affected by previous schooling. Indeed, Meghir et al. (2011), using the same Swedish cognitive test score data as Sandewall et al. (2009), find that this is the case.[4] Nevertheless, it is important to note that even if twins would differ in ability, the twin design is still useful in tightening the upper bound of the estimated effects (Bound and Solon, 1999). This is also the view we take in this article; our WTP estimates should be regarded as an upper bound of the effect of income, and other socioeconomic factors, on health. In this sense, they will indicate

---

[3] Note, however, that parents may respond by compensating for differences in for instance health or ability among their twins, which may partly offset any upward bias. Some evidence for such behavior is presented in Lundborg (2012), where the twin who reported greater time investment by parents during childhood also reported less education on average, suggesting that parents invested more in the "weaker" twin. However, other twin-based studies, using various measures of parental inputs, do not suggest that parents systematically reinforce or compensate for early life insult (see for instance Royer, 2009; Almond and Currie, 2011). Moreover, Isacsson (1999) finds no relation between psychological instability early in life (an imperfect proxy of parental rearing skills) and years of schooling among Swedish twins. We are not aware of any evidence suggesting that parents systematically reinforce existing differences, which would result in upward biased estimates.

[4] They show that the Swedish compulsory schooling reform had a significant impact on the development of cognitive and non-cognitive abilities of young men during their teens.

the minimum extent to which OLS estimates on the relation between socioeconomic factors and health are biased and hence inappropriate to use as a decision basis.

The second potential problem with a twin design is the well-known fact that the importance of normally distributed measurement errors in explanatory variables is exacerbated by differencing, and even more so when differencing between MZ twins (Griliches 1979). This may cause twin fixed effects estimates of our explanatory variables to be downward biased. As shown by Griliches (1979), in the presence of classical measurement error, the WTP estimates would be biased according to:

$$\beta_{WTP} = (1 - ((Var(v))/([Var(y)](1 - \rho y)))), \qquad (11)$$

where $Var(v)$ denotes the assumed common variance of the twins measurement error, $Var(y)$ is the variance in the true income levels, and ρy is the correlation between the measured income of the twins. Here, we expect the measurement error problems in the WTP estimation to be relatively small, since register information is used for the key factors of income and education. Holmlund et al. (2008) found high reliability for Swedish register-based measures of education; 0.95 for males and 0.94 for females. If the reliability ratios are similar for income, which one may expect since income is also register-based, the measurement error problems would not severely threaten our estimates.

For variables that are based on dummy variables, such as employment status, measurement errors will be non-classical. The reason for this is that individuals in the lowest category cannot under-report their employment status, whereas individuals in the highest category cannot over-report (Aigner, 1973). With non-classical measurement error, one cannot generally sign the bias in the estimates.

3. Sample and variable definitions

Sample construction

This article exploits micro data originating from the Swedish Twin Registry. The study population consists of a subset of the data set; the participants in a telephone interview called Screening Across the Lifespan Twin Study (SALT). This is the only part of the twin register where data on self-reported health is available, which is our main dependent variable. The Swedish Twin Registry was established in the 1950s to study the health consequences of smoking and alcohol consumption, and is now the largest population-based twin register in the world. It has been described in detail previously (Lichtenstein et al., 2002, 2006). The SALT survey was conducted in 1998–2002, and attempted to include all Swedish twins born in 1958 or earlier, and hence aged 40 years or older at the time of the interview. Thus the study includes the population of twins born in 1896–1958 who survived until 1998-2002, which provides us with 44,919 observations including 31,090 complete twin pair observations; that is, 15,545 twin pairs (11,357 DZ, 4,079 MZ, and 109 with unknown DZ-MZ characteristic). The survey data include detailed data on a wide range of health outcomes, labor market and civil status variables, and more. Moreover, the data are matched with register information on income and education. The income data was also recorded retrospectively prior to the interview, and education information was taken from the years 1990 and 2007.

Summary statistics for the samples of the twins are included in Table 1, both for the total number of observations available for each variable in different twin samples and for the observations of complete pairs that are used in the analysis. Data from the Statistics Sweden Survey of Living Conditions (the ULF survey) are also used to complement the twin data, for two reasons: 1) to cardinalize our self-reported health variable (see below), and 2) to explore the representativeness of the twin data. The ULF survey data are based on yearly one-hour personal interviews with about 6,000 randomly selected adults in Sweden aged 16-84 years, and include many variables similar to those in the twin data set. In the cardinalization

exercise, we use ULF data from the 2004/2005 survey, where variables are available for EQ-5D-like dimensions of health. In investigating the representativeness of our twin data, we use pooled data from the 1998-2002 ULF surveys which match the years used in the twin data set.

Variables

In order to facilitate comparison with previous studies based on cross-sectional OLS-based decomposition analyses, we use variables similar to the ones used in previous studies. Thus, we use self-reported health as our dependent health variable, and include gender, age, income, education, employment, civil status, and number of children as independent variables in the decomposition analysis.

Self-reported health

The self-reported health measure is a multiple-category indicator in which individuals answer the question 'How is your health in general?' The answer is chosen from 5 categories: very poor (2.4%), poor (8%), fair (21%), good (36%), and very good (33%). An important aspect of twin data analyses is the presence of sufficient variation within twin pairs. In total, 5,410 (36%) twin pairs report equal health, of which 1,668 (41%) are MZ.

Since the self-reported health information is not continuous but measured on an ordinal scale, the concentration index approach cannot be applied in a direct manner – the problem being that the health of individuals cannot be aggregated in a meaningful way. The concentration index approach requires a health variable that is measured at least on a cardinal scale. There are several ways to deal with this problem. This study performs a cardinalization procedure in three steps. The first step creates a link between EQ5D-like health state scores and answers to self-reported health question of the kind that the respondents in our data set also answered. For this purpose, we use the ULF survey data from 2004/2005, covering a

random sample of the Swedish population aged 16-84 years (individuals aged 15-39 years are omitted here in order to cover the same ages as in the twin data set). Using the algorithm in Burström et al. (2001, 2003, 2005) and Islam et al. (2010), the ULF respondents can be classified along lines which are close to the EQ-5D-dimensions (Dolan, 1997). Health scores for these profiles are then taken from the UK EQ-5D index tariff, since there is no Swedish time trade-off tariff for EQ-5D health states (Dolan, 1997). The mean value of the estimated EQ-5D health state scores in the ULF data set is 0.766 (Std. Dev. =0.3074), ranging from 0 to 1. In the second step, these health scores are regressed on gender, age, and categorical self-assessed health (with very poor health as baseline), giving the following regression results (all coefficients estimated are significant at the 1% level):

$$h_i = 0.274 + 0.045 \cdot Male_i - 0.001 \cdot Age_i + 0.716 \cdot VeryGood \sim Health_i +$$
$$0.625 \cdot Good \sim Health2_i + 0.384 \cdot Fair \sim Health_i + 0.066 \cdot Poor \sim Health_i$$

Number of observations: 5,844. R^2=0.44

Finally, in the third step, these estimates are used to predict EQ-5D-like health state scores in the twin data set, giving us a cardinal health variable to use in the analysis. The mean health score in the entire complete pair twin sample is 0.778 (Std. Dev.=0.206), ranging from 0.171 to 0.984, and for the MZ sample it is 0.782 (Std. Dev.=0.203), ranging from 0.178 to 0.983. For the representativity analysis of the twin data, the above estimated function based on data for 2004/2005 is used to predict the health scores in the ULF survey data for 1998-2002, which matches the years in the twin data.

Income

Income is measured as individual earned taxable income, with the consumer price index used to deflate income across the years to the price level of 2003.[5] Annual income (and also education; see below) is expected to carry small measurement errors, since income data is obtained from registers. However, it has been suggested that "lifetime" or permanent income may be more appropriate to use in an analysis of health behavior (Frijters et al. 2005). If that is the case, our income measure will still include measurement errors due to transitory income. To reduce this problem, a 10-year average of annual income within individuals is used in the analysis.

To give an impression of how much the estimated average income differs within twin pairs, we divide the income variable into quintiles of the average income distribution and then calculate the proportion of twin pairs where both siblings are in the same income quintile. This procedure shows that 4,974 (36%) of the complete twin pairs and 1,835 (45%) of the MZ twin pairs are in the same income quintile.

Education

Our education variable is constructed from register data and measured according to SUN (Swedish Educational Terminology), the standard system for classifying education in Sweden. The data contain years 8-20 of schooling. According to the register information, 36% of all twins and 47% of MZ twins have the same level of education.

Other socioeconomic variables

---

[5] It is not possible to calculate after-tax household income or equivalent after-tax household income, since only before-tax income is available in the data and the income of any other household members is unknown.

At the individual level, other factors potentially affecting health include age, gender, and civil status (i.e. whether an individual is divorced, widowed, unmarried, or married, the last of these being the baseline). Number of children in the household is measured by inclusion of three indicator variables capturing whether there are one, two, or three or more children in the household. Economic activity status is measured by the following seven categories: working part-time, self-employed, unemployed, house working, student, economically inactive (early retirement, disability pensioner, long term absenteeism due to sickness), and retired. The baseline category is employed. The variation within twin pairs in regard of economic activity status variables is low, with 79% of all twin pairs and 82% of MZ twins reporting equal activity status.

Table 2 displays the estimated mean health in the different deciles of the income distribution for 1) all twins, 2) all complete twin pairs, 3) complete MZ twins, and 4) complete MZ twins with income deciles based on annual income. In 5) the corresponding mean health is shown across annual income based on the ULF data set. In addition, the C's of health scores are calculated for each sample. Mean health is similar among the different twin samples, columns 1-4. As expected, mean health is slightly higher for complete twin pairs (2-3) compared with samples based on all twins (1), but the differences are small. The C of health when individuals are ranked after annual income is slightly higher than when ranked by 10-year average income.[6] The C of health for the "representative" sample of the Swedish population is somewhat higher compared with the MZ sample; 0.052 versus 0.047.

---

[6] It is important to note however that C of ill-health (1-health) would not be equal to the negative of C of health, since C does not fulfill the mirror condition discussed in the recent literature on the properties of C (Erreygers, 2009a; Erreygers and van Ourti, 2010). Erreygers (2009a) developed a correction of the concentration index which also fulfills the mirror condition. In the present article, however, we use the conventional C, since the focus is

4. Estimation results

This section describes the results of the empirical analysis, and is organized as follows. First, Table 3 presents the estimates of the health production function for different twin samples, based on OLS and WTP analysis. Second, Table 4 presents the results of the OLS and WTP decomposition analyses, respectively.[7] For each of the health factors, we report health elasticities, C's, and contributions expressed in absolute terms (Eff = elasticity multiplied by the corresponding C) and in percent (Eff% = Eff/(health C)*100). The measure Agg% sums the percentage contributions over the categorical variables such as age and labor market status (e.g., in panel 1, summing Eff% from "working part-time" to "retired" gives the value 19.51). Third, Table 5 presents an additional decomposition analysis based on OLS and data from the twin and ULF samples, to illustrate the representativeness of the twin data set.

OLS and WTP estimation of health production (Table 3)

OLS estimations based on complete twin pairs are reported in columns 1-3 of Table 3, and the corresponding WTP estimations are reported in columns 4-6. Columns 1 and 4 represent DZ and MZ twins combined, columns 2 and 5 represent DZ twins, and columns 3 and 6 represent MZ twins. The estimates based on OLS are in line with past studies, and most effects are significant at least on the 5% level in the "expected" direction. For example, income and education have protective effects and are highly significant. Most labor market and civil status variables are also significant.

---

mainly on the percentage contributions of the health factors on the concentration index, which will be identical irrespective of which C is used for the decomposition analysis.

[7] The OLS-based and WTP-based decomposition analyses for DZ twins are available on request.

The question now is whether these results can be interpreted causally, giving the conclusion that people who invest more in education or achieve a higher income will also experience health improvements. Conversely, it could be the case that people with higher innate ability and advantageous family background are able to obtain both better health and better income and education. In the latter case, a causal relationship between health and socioeconomic status is not necessarily implied. Our results suggest strongly that the latter hypothesis is closer to the truth, since the results change radically when we move to the WTP estimation. The effects of education and income are now substantially reduced. In contrast, the effects of the labor market and civil status factors seem even somewhat stronger in the WTP estimation.

The main concern of this article is however not the effects of socioeconomic factors on health (though this is obviously a crucial part of the analysis), but more specifically the contribution of these socioeconomic factors to measured income-related inequalities in health. Thus the next step estimates and compares the OLS-based and WTP-based decomposition analyses.

Decomposition of the inequality-health index into its constituent parts

This section presents the results of the decomposition in three subsections. The first subsection presents C of health and C of the factors in the health function, while the second and third subsections report the contributions of the various factors on C of health based on OLS and WTP estimates, respectively. All these results are given in Table 4.

a) Estimation of concentration indexes

C of health is estimated to be around 0.04 in the full sample, as well as in the sample restricted to MZ twins, indicating that (as expected) the health variable is distributed to the

advantage of people with higher incomes. For the determinants of health, *C*'s are robust across the samples and indicate that males and people aged 51-60 are concentrated to higher levels of income, while older individuals are concentrated to lower levels of income. *C*'s of (log of) income and education are positive, indicating that people with higher income (per definition) and education are more frequently distributed in higher income ranks (pro-rich). People who are working part-time, unemployed, or economically inactive are distributed towards lower income (pro-poor). The distributions of self-employed people, retired people, and those with children are pro-rich, while the distributions of widowed and unmarried people are pro-poor. Many of these results are self-evident, and hence serve as a credibility check.

b) OLS-based health factor contributions

As shown in Table 4 (panel 1, columns 1-5) income and education contribute 24% and 10%, respectively, to *C* of health for all complete twin pairs. Income is the single most important contributor to income-related inequalities in health. The labor market variables contribute 20% to inequality. In total, 33% of inequalities are explained by gender and age and the rest are explained by civil status, number of children in family, and a small residual term. The results are similar for all complete MZ twins (panel 3, columns 11-16). In sum, these results are qualitatively similar to those found in earlier studies; that is, income (in particular), education, and labor market variables (especially economic inactivity) are important contributors to *C* of health.

c) WTP-based health factor contributions

The WTP-based results for all twin pairs reported in Table 4 (panel 2, columns 6-10) show that the contributions of income and education are substantially reduced compared to the OLS-based analysis above, but are still significant contributors to *C* of health: income and

education contribute 17% and 5%, respectively. Several labor market variables also contribute significantly, with similar effects as in the OLS-based analysis (21%). However, the dominant contributor to C of health is the estimated fixed effects, which contribute 35% (bottom line in the table). Generally, these fixed effects capture contributions of all factors common to the twin pair; observed factors such as age, as well as unobserved genetic predispositions and common family background. Since the contribution of the fixed effects is twice the contribution of age in the OLS analysis, we conclude that age, known to be intimately connected with health, does not explain the whole fixed effect.

In the WTP analysis based exclusively on MZ twins (panel 4, columns 16-20), the contribution of income is further reduced to 11% and the contributions of income and education (still about 5%) are now both insignificant. Note that these contributions could still be regarded as upper bounds of the true contributions if there remains important unobserved heterogeneity within twin pairs relating to both socioeconomic factors and health. In any case, the role for income and education policies in affecting health inequalities appears much more limited than that suggested by the OLS estimates. The contributions of labor market variables are roughly the same as in the OLS-based analysis (19%). The main contributor to C of health is unquestionably the fixed effects; 64% of the income-related inequalities in health are explained by the twin pair fixed effects. This increase from the figure of 35% seen for all twins may reflect the identical genetic structure of MZ twins compared with DZ twins, suggesting that genetics is an important third factor behind the relationship between socioeconomic status and health.

Figure 1 summarizes Table 4. It appears from the figure that the contributions from income and education are reduced when the health production function is estimated by WTP, while the contribution from the fixed effects increases. It is also clear that the contribution

from the fixed effects increases further when the estimating sample consists of only MZ twins.

To illustrate the representativity of the twin data in relation to the general Swedish population of the same age range, the decomposition results in Table 4 are reproduced in Table 5 based on "restricted" OLS regressions including only gender, age, annual income, and schooling years; that is, variables which are identically defined in the twin and ULF samples. The contributions of income and education amount to about 36% and 14% respectively in the MZ twin sample and 38% and 13% in the ULF sample, indicating a high degree of representativity among the analyzed twins.

5. Discussion

Earlier economic analyses on the explanation of income-related inequalities in health using the concentration index and decomposition analysis have universally demonstrated that income and education are key contributors to the observed health inequalities as measured by the concentration index. The policy relevance of these results may appear apparent from an efficiency perspective, since they indicate where decision makers should allocate resources in order to achieve as much as possible in terms of reduced inequalities in health. These policy conclusions are also in line with the Marmot Review (Marmot et al., 2010), which argues that health inequalities are the results of socioeconomic inequalities and that socioeconomic inequalities (e.g. in income) need to be reduced in order to reduce socioeconomic health inequalities.

Such conclusions are based on the assumption that socioeconomic factors have a causal effect on health, however, and this remains in dispute. As is well known, the association between income and education, on the one hand, and health, on the other, may result from a variety of sources. Unobserved genetic and family background traits that are

23

correlated with income and education as well as health status would induce a positive bias in the estimation of the health production function. In earlier decomposition analyses, the endogeneity issue of the health factors is essentially overlooked (which is understandable in view of the difficulties involved). Most analyses have been based on OLS using cross-sectional data from which it is virtually impossible to identify causal effects. As a consequence, policies based on such analyses risk being ineffective if they target factors which in reality are only related to health through underlying factors such as family background or genetic predispositions.

The current article uses the twin design in order to remove the influence of many of the abovementioned unobserved underlying factors. The WTP-based analysis demonstrates that the contributions of income and education are significantly reduced compared with the OLS-based analysis. In addition, the WTP-based analysis shows that the main contributor to the income-related inequalities measure is the WTP fixed effect, which incorporates the influence of all factors shared by the twins. Besides unobserved genetic and family background, this also captures factors such as age and gender. In total, the fixed effect explains 64% of the estimated health concentration index. Our results may serve as an indication of the limit to what decision makers can do about inequalities in health.

In summary: our paper provides a number of key insights which are highlighted below. First, in prior studies, where the decomposition analysis is based on OLS estimation of the health production function using cross-sectional data, the estimated contribution of income and education is probably substantially exaggerated. Second, the WTP-based analysis gives more rigorous estimates and provides an upper bound for the contribution of the health factors to the inequalities in health. The results are obviously important for policy makers, but leave them somewhat in limbo. The effects of income and education, which could be prime targets for policy interventions, are insignificant. If anything, among the observed

socioeconomic factors, it is labor market variables (more specifically, the state of being economically inactive) which contribute most to the inequalities in health. The main part of the inequalities in health are instead explained by the WTP fixed effect, which leaves the decision makers without much guidance since this fixed effect encapsulates a variety of potential factors, some of which might be possible to manipulate while others are not. In any case, none of these factors have yet been identified. It should also be noted that this study concerns self-reported overall health, and it is possible that certain specific health outcomes are more directly influenced by the socioeconomic status of individuals.

In the long run, we need to find out more about the hard-to-measure factors that affect health as well as education and/or income. We have highlighted the fact that income and education are insignificant when we restrict our sample to MZ twins. It is worth pointing out that this still leaves a potential scope for important contributions of income differences in society. As noted above, one factor that is shared by twins is the relative affluence of their family as they grow up. In other words, it is certainly possible that their parents' incomes and education have a significant influence on their health as children as well as later in life, and also that the parents' incomes and education affect the children's incomes and education.[8] If so, reducing income differences may in fact still be a way to reduce socioeconomic inequalities in health, but it may take a generation to see any results.

---

[8] For recent twin-based evidence of this, see for instance Lundborg et al. (2011), Amin et al. (2011a), and Amin et al. (2011b).

References

Aigner, D. 1973. Regression with a binary independent variable subject to errors of observation. Journal of Econometrics 1, 49-60.

Almond, D, Currie, J. (2011). "Human Capital Development before Age Five," Handbook of Labor Economics, David Card and Orley Ashenfelter (eds.).

Amin, V, Behrman, J. 2009. Does more schooling improve health behaviours and health outcomes? Evidence from U.K twins. Unpublished manuscript.

Amin, V, Lundborg, P, Rooth DO. (2011a) Following in Your Father's Footsteps: A Note on the Intergenerational Transmission of Income between Twin Fathers and their Sons IZA Working Paper 5990.

Amin, V, Lundborg, P, Rooth DO. (2011b) Mothers Do Matter: New Evidence on the Effect of Parents' Schooling on Children's Schooling Using Swedish Twin Data. IZA Working Paper 5946.

Ashenfelter, O, Krueger, A. 1994. Estimating the returns to schooling using a new sample of twins." American Economic Review, 84(5): 1157-1174.

Ashenfelter, O, Rouse, C. 1998. Income, schooling and ability: evidence from a new sample of identical twins. Quarterly Journal of Economics, 113(1), 253.284.

Baum C, Ruhm C. 2009. Age, socioeconomic status, and obesity growth, Journal of Health Economics, 28: 635-648.

Behrman, J.R,.Rosenzweig, M.R. 2004. Returns to birthweight. Review of Economics and Statistics 86, 586-601.

Behrman, J, Kohler, H.P, Jensen, V.M, Petersen, I, Pedersen, D, Bingley, P, Christensen, K. 2011: Does more schooling reduce hospitalization and delay mortality? Demography 47(3)

Benzeval, M, Judge, K. 2001. Income and health: the time dimension. Social Science & Medicine 52, 1371–1390.

Black, S.E., Devereux, P.J, Salvanes, K.G. 2007. From the cradle to the labor market? The effect of birth weight on adult outcomes. Quarterly Journal of Economics, 122(1): 409-439.

Bloom, D.E, Canning, D. 2000. `The health and wealth of nations', Science 287(5456), 1207-1209.

Bonjour, D, Cherkas, L, Haskel, J, Hawkes, D, Spector, T. 2003. Returns to Education: Evidence from UK Twins. American Economic Review, 93(5): 1799-1812.

Bound, J, Solon, G. 1999. Double trouble: on the value of twin-based estimation of the return to schooling. Economics of Education Review 18, 169-182.

Buckley, N.J, Denton, F.T, Leslie R, Spencer, B.G. 2004. The transition from good to poor health: an econometric study of the older population, Journal of Health Economics, Elsevier, vol. 23(5), pages 1013-1034, September.

Burström K, Johannesson M, Diderichsen F. 2001. Health-related quality of life by disease and socio-economic group in the general population in Sweden. *Health Policy*, **55**: 51-69.

Burström K, Johannesson M, Diderichsen F. 2003. The value of the change in health in Sweden 1980/81 to 1996/97. *Health Economics,* **12**: 637-654.

Burström K, Johannesson M, Diderichsen F. 2005. Increasing socio-economic inequalities in life expectancy and QALYs in Sweden 1980 to 1997. *Health Economics*, **14**: 831-850.

Costa-Font, J, Fabbri, D, Gil, J. 2009. Decomposing body mass index gaps between mediterranean countries: a counterfactual quantile regression analysis. Economics of Economics & Human Biology, Vol. 7, No. 3, pp. 351-365, 2009.

Cunning, D. and Bowser, D. 2010. Investing in Health to Improve the Wellbeing of the Disadvantaged: Reversing the Augment of the Marmot Reports. PGDA Working Paper No. 78 http://www.hsph.harvard.edu/pgda/working.htm. May 2010.

Cutler D, Lleras-Muney A, Vogl T. 2008. Socioeconomic Status and Health: Dimensions and Mechanisms, NBER Working Paper No. 14333.

Deaton A. 2002. Policy implications of the gradient of health and wealth an economist asks, would redistributing income improve population health?, Health Affairs, 21, 13-30.

Deaton, A.S. 2003. Health, inequality, and economic development. Journal of Economic Literature XLI, 113–158.

Deaton, A. 2011. What does the empirical evidence tell us about the injustice of health inequalities. Center for Health and Wellbeing, Princeton University, January.

Dolan P. 1997. Modeling valuation for EuroQoL health states. *Medical Care*, **35**: 1095-1098.

Erreygers, G. 2009a. Correcting the concentration index. Journal of Health Economics 28, 504-515.

Erreygers, G. 2009b. Correcting the concentration index: a reply to Wagstaff. Journal of Health Economics 28, 521-524.

Erreygers, G, van Ourti, T. 2010. Measuring socioeconomic inequality in health, health care and health financing by means of rank dependent indices: a recipe for good practice" Tinbergen Institute Discussion Papers 10-076/3, Tinbergen Institute.

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

Fleurbaey, M, Schokkaert, E. 2011. Equity in health and health care, CORE Discussion Papers 2011026, Université catholique de Louvain, Center for Operations Research and Econometrics.

Frijters, P, Haisken-DeNew, J.P, Shields, M.A. 2005. The causal effect of income on health: Evidence from German reunification, Journal of Health Economics, Elsevier, vol. 24(5), pages 997-1017, September.

Fuchs, V.R. 1982. Time Preference and Health: An Exploratory Study", in V. Fuchs (ed.) *Economic Aspects of Health* (Chicago: The University of Chicago Press).

Gerdtham U-G, Johannesson M. 2000. Income-Related Inequality in Life-Years and Quality-Adjusted Life-Years in Sweden, Journal of Health Economics, **19**: 1007-1026.

Gerdtham U-G, Johannesson M. 2002. Do Life-Saving Regulations Save Lives?, Journal of Risk and Uncertainty, **24**: 231-249.

Gerdtham U-G, Johannesson M. 2004. Absolute Income, Relative Income, Income Inequality and Mortality?, Journal of Human Resources, **39**: 228-247.

Gomez, P.M., Lopez-Nicholas, A. 2005. Socio-economic inequalities in health in Catalonia. *Hacienda Pública Española / Revista de Economía Pública,* 175-(4/2005): 103-121

Gravelle, H., Sutton, M., 2003. Income related inequalities in self assessed health in Britain: 1979–1995. Journal of Epidemiology and Community Health 57, 125–129.

Griliches, Z. 1979. Sibling models and data in economics: beginnings of a survey. Journal of Political Economy 87(5): 37.64.

Holmlund, H, Lindahl, M. Plug, E. 2008. The causal effect of parents' schooling on children's schooling: a comparison of estimation methods", IZA Discussion Paper No. 3630.

Hosseinpoor, A.R, Van Doorslaer, E.V, Speybroeck,N, Naghavi,M.,Kazem M., Majdzadeh, R., Delavar, B., Jamshidi, H.,Vega, J. 2006. Decomposing socioeconomic inequality in infant mortality in Iran. International Journal of Epidemiology 2006,35:1211–1219.

Isacsson, Gunnar (1999), "Estimates of the Return to Schooling in Sweden from a Large Sample of Twins", Labour Economics 6:4, pp. 471-489

Isacsson, G. 2004. Estimating the economic returns to educational levels using data on twins. Journal of Applied Econometrics 99: 99-119.

Islam, M, Gerdtham, U.-G, Clarke, P, Burstrom, K. 2010. Does income-related health inequality change as the population ages? Evidence from Swedish panel data. Health Economics 19 (3), 334–349.

Jones, A.M. Lopez Nicolas, A. 2006. Allowing for heterogeneity in the decomposition of measures of inequality in health. The Journal of Economic Inequality 4: 347-365DOI: 10.1007/s10888-005-9019-z.

Jones, A.M. Schurer, S. How does heterogeneity shape the socioeconomic gradient in health satisfaction. Ruhr Economic Papers #8. Ruhr Graduate School in Economics. Germany.

Kakwani N.C, Wagstaff A, van Doorslaer E. Socioeconomic inequalities in health: measurement, computation and statistical inference. Journal of Econometrics 1997; 77(1): 87–104.

Kjellsson, G, Gerdtham, U-G. 2011. Correcting the concentration index for binary variables. Working Paper, Department of Economics, Lund University No 2011:4.

Kunst, A, Bos, V, Lahelma, E, Bartley, M, Lissau, I, Regidor, E, Mielck, A, Cardano, M, Dalstra, J, Geurts, J, Helmert, U, Lennartsson, C, Ramm, J, Spadea, T, Stonegger, W, Mackenbach, J. 2005. Trends in socio-economic inequalities in self-assessed health in 10 European countries. International Journal of Epidemiology 34, 295–305

Lauridsen J, Christiansen T, Gundgaard J, Häkkinen U, Sintonen H. 2007. Decomposition of health inequality by determinants and dimensions. Health Economics, Jan;16(1):97-102.

Leu, R.E, Schellhorn, M. 2004. The Evolution of Income-Related Health Inequalities in Switzerland over Time. IZA DIscussion Paper Series No. 1346.

Lichtenstein, P, de Faire, U, Floderus, B, Svartengren, M, Svedberg, P, Pedersen, N.L. 2002. The Swedish Twin Registry: A Unique Resource for Clinical, Epidemiological and Genetic Studies. Journal of Internal Medicine, 252(3): 184-205.

Lichtenstein, P, Sullivan, P.F, Cnattingius, S, Gatz, M, Johansson, S, Carlström, E, Björk, C, Svartengren, M, Volk, A, Klareskog, L, de Faire, U, Schalling, M, Palmgren, J, Pedersen, N.L. 2006. The Swedish Twin Registry in the Third Millennium: An Update Twin Research and Human Genetics, 9(6): 875-882.

Lindahl, M. 2005. Estimating the effect of income on healthand mortality using lottery prizes as an exogenous source of variation in income. The Journal of Human Resources 40, 144-168.

Lundborg P. (2012).The health returns to education –what can we learn from twins? Forthcoming in Journal of Population Economics.Lundborg, Petter, Martin Nordin, and Dan-Olof Rooth. 2011b. "The Intergenerational Transmission of Human Capital: Exploring the Role of Skills and Health Using Data on Adoptees and Twins." Institute for the Study of Labor Discussion Paper 6099.

Mackenbach, J.P, Bos, V, Andersen, O, Cardano, M, Costa, G, Harding, S, Reid, A, Hemström, Ö, Valkonen, T, Kunst, A.E.. 2003. Widening socioeconomic inequalities in mortality in six Western European countries, International Journal of Epidemiology, 32 (5),830-837.

Mackenbach, J.P. 2005. Genetics and health inequalities: hypotheses and controversies. 2004. Journal of Epidemiology & Community Health 2005;59;268-273

Marmot M, Allen J, Goldblatt P, et al. 2010. Fair society, healthy lives: the Marmot review – strategic review of health inequalities in England post–2010. London: University College London.

Marmot, M. 2012. Policy Making With Health Equity at Its Heart. JAMA, May 16, 307, 2033-2034.

McGrail, K.M, van Doorslaer, E, Ross, N.A, Sanmartin, C. 2009. Income-related health inequalities in Canada and the United States: A decomposition analysis. American Journal of Public Health 99: 1856-1863.

Meghir, C., Palme, M., Simeonova, E. (2011). How Does Education Affect Labor Market Outcomes: The Effect of Schooling on Cognitive and Non-Cognitive Skills. Working Paper.

Meer, J, Miller, D.L, Rosen, H.S. 2003. Exploring the health–wealth nexus. Journal of Health Economics 22, 713-730.

Meghir, C, Palme, M, Simeonova, E. 2012. Education, Health and Mortality: Evidence from a Social Experiment. NBER Working Paper No. 17932 March 2012

Miller, P, Mulvey, c, Martin, N. 2005. "Birth Weight and Schooling and Earnings: Estimates from a Sample of Twins." Economics Letters 86(3): 387--392.

Morasae, E.K, Forouzan, A.S, Majdzadeh, R, Asadi-Lari, M, Noorbala, A.A, Hosseinpoor, A.R. 2012. Understanding determinants of socioeconomic inequality in mental health in Iran's capital, Tehran: a concentration index decomposition approach. International Journal for Equity in Health 11:18 http://www.equityhealthj.com/content.

O'Donnell, O, van Doorslaer, E, Wagstaff, A, Lindelow, M. 2008. *Analyzing health equity using household survey data : a guide to techniques and their implementation*. Washington D.C.: World Bank.

Petersen I, Jensen VM, McGue M, Bingley P, Christensen K. 2009. "No Evidence of Genetic Mediation in the Association Between Birthweight and Academic Performance in 2,413 Danish Adolescent Twin Pairs." Twin Research and Human Genetics 12(6):564-72.

Royer, H. 2009. Separated at girth: US twin estimates of the effects of birth weight." American Economic Journal: Applied Economics, 1(1): 49--85.

Sandewall, Ö, Cesarini, D, Johannesson, M. 2009, The co-twin methodology and returns to schooling – testing a critical assumption", IFN Working Paper no. 806, Research Institute of Industrial Economics (IFN).

Shkolnikov V.M, Andreev E.M, Jdanov D.A, Jasilionis D, Kravdal O, Vagero D, Valkonen T. 2011. Increasing absolute mortality disparities by education in Finland, Norway and Sweden, 1971-2000. Journal of Epidemiology & Community Health 66; 372–378.

Smith J. 1999. Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status, Journal of Economic Perspectives, **13**: 145-166.

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

Vasoontara, Y, Lim, L.L.Y., Carmichael, G.A.,Dear, K.B.G., Sleigh, A.C. 2010. Decomposing socioeconomic inequality for binary health outcomes: an improved estimation that does not vary by choice of reference group. BMC Research Notes 3:57.

Wagstaff, A., Van Doorslaer, E., 2000. Equity in health care finance and delivery. In: Culyer, A.J., Newhouse, J.P. (Eds.), Handbook of Health Economics. Amsterdam, Netherlands, North Holland, pp. 1803–1862.

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

Webbink, D, Nicholas G.M, Visscher, P.M. 2010. Does education reduce the probability of being overweight? Journal of Health Economics 29, Issue: 29-38.

Wildman, J. 2003. Modelling health, income and income inequality: the impact of income inequality on health and health inequality. Journal of Health Economics, Elsevier, vol. 22(4), pages 521-538, July.

Table 1: Means of basic samples and working samples of the Swedish Twin Registry (SALT) and the ULF survey (different years).

	(1)	(2)	(3)	(4)	(5)
	All Twins n=44,919 – 43,373	Twins included in the complete twin pairs regressions n=31,090[*]	Twins comprising complete DZ twin pairs n=22,714	Twins comprising complete MZ twin pairs n=8,158	ULF pooled 2004-2005 sample n=16,481
Variable					
Health: 1 (excellent)	0.325	0.341	0.338	0.348	0.302
Health: 2	0.363	0.368	0.367	0.370	0.377
Health: 3	0.208	0.197	0.199	0.192	0.239
Health: 4	0.080	0.074	0.075	0.071	0.063
Health: 5 (bad)	0.024	0.020	0.021	0.019	0.020
Male	0.465	0.464	0.474	0.436	0.479
Age	59.342	57.377	57.434	57.198	59.022
Average income[**]	203,048	212,818	211,988	215,449	221,996
(Annual income)[**]	(220,137)	(234,221)	(230,132)	(237,236)	(221,996)
Schooling years	10.953	11.124	11.074	11.267	11.043
Working part-time	0.124	0.139	0.140	0.138	-
Self-employed	0.055	0.062	0.062	0.061	-
Unemployed	0.020	0.021	0.021	0.019	-
Economically inactive	0.085	0.090	0.089	0.091	-
Retired	0.020	0.022	0.023	0.021	-
Divorced	0.089	0.093	0.090	0.099	-
Widowed	0.084	0.064	0.064	0.064	-
Unmarried	0.098	0.095	0.096	0.091	-
Children: 1	0.090	0.100	0.099	0.104	-
Children: 2	0.063	0.070	0.069	0.072	-
Children: ≥3	0.025	0.028	0.028	0.028	-

[*]Including 218/2=109 twin pairs with unknown DZ-MZ characteristics.
[**]Expressed in 2003 prices.

Table 2: Means of health scores in different income deciles from low to high in the twin register and ULF data.

Deciles	All twins	Complete twins	MZ twins (complete)	MZ twins (complete)[*]	ULF data[*]
1	0.665	0.681	0.681	0.682	0.676
2	0.683	0.702	0.709	0.697	0.683
3	0.712	0.737	0.745	0.726	0.702
4	0.734	0.748	0.757	0.750	0.729
5	0.753	0.771	0.779	0.772	0.747
6	0.780	0.790	0.795	0.799	0.791
7	0.796	0.807	0.799	0.824	0.808
8	0.816	0.823	0.824	0.831	0.840
9	0.842	0.849	0.846	0.852	0.850
10	0.872	0.877	0.884	0.888	0.879
N	43,651	31,090	8,158	8,072	16,481
Mean health	0.770	0.778	0.782	0.782	0.770
C of health	0.048	0.044	0.043	0.047	0.052
SE	0.0007	0.0009	0.0016	0.0016	0.0011

[*]Income deciles based on annual income.

Table 3: Pooled and WTP regression estimations of the determinants of health (SE are robust standard errors)[),*]

Variables	Complete twin pairs											
	OLS estimation						**WTP estimation**					
	(1) DZ+MZ		**(2) DZ**		**(3) MZ**		**(4) DZ+MZ**		**(5) DZ**		**(6) MZ**	
	Coeff	SE	Coeff	SE	Coeff*	SE	Coeff	SE*	Coeff*	SE	Coeff*	SE
Male	0.044[a]	0.0024	0.045[a]	0.0028	0.045[a]	0.0048	0.045[a]	0.0037	0.044[a]	0.0038		
Age 51-60	-0.019[a]	0.0028	-0.021[a]	0.0033	-0.016[a]	0.0052						
Age 61-70	-0.038[a]	0.0037	-0.038[a]	0.0043	-0.034[a]	0.0072						
Age 71-84	-0.118[a]	0.0048	-0.119[a]	0.0055	-0.112[a]	0.0093						
L(Income)	0.029[a]	0.0024	0.028[a]	0.0028	0.033[a]	0.0048	0.021[a]	0.0034	0.022[a]	0.0038	0.013[c]	0.0076
Schooling	0.006[a]	0.0004	0.006[a]	0.0005	0.005[a]	0.0009	0.003[a]	0.0008	0.003[a]	0.0009	0.003	0.0017
Working part-time	-0.031[a]	0.0034	-0.032[a]	0.0040	-0.024[a]	0.0065	-0.042[a]	0.0046	-0.041[a]	0.0054	-0.038[a]	0.0085
Self-employed	0.031[a]	0.0036	0.032[a]	0.0042	0.029[a]	0.0071	0.024[a]	0.0056	0.020[a]	0.0063	0.031[a]	0.0117
Unemployed	-0.029[a]	0.0082	-0.025[a]	0.0090	-0.035[c]	0.0183	-0.053[a]	0.0105	-0.044[a]	0.0119	-0.070[a]	0.0218
Economically inactive	-0.253[a]	0.0051	-0.251[a]	0.0060	-0.256[a]	0.0100	-0.255[a]	0.0068	-0.265[a]	0.0079	-0.224[a]	0.0137
Retired	0.004	0.0067	0.004	0.0079	0.005	0.0127	-0.049[a]	0.0114	-0.042[a]	0.0133	-0.070[a]	0.0211
Divorced	-0.030[a]	0.0039	-0.028[a]	0.0046	-0.034[a]	0.0075	-0.020[a]	0.0053	-0.022[a]	0.0063	-0.013	0.0096
Widowed	-0.039[a]	0.0053	-0.039[a]	0.0062	-0.039[a]	0.0101	-0.014[b]	0.0072	-0.018[b]	0.0087	-0.002	0.0123
Unmarried	-0.035[a]	0.0039	-0.040[a]	0.0046	-0.021[a]	0.0073	-0.040[a]	0.0053	-0.041[a]	0.0061	-0.036[a]	0.0106
Children: 1	0.006[c]	0.0034	0.009[b]	0.0040	-0.003	0.0065	-0.002	0.0047	0.006	0.0057	-0.025[a]	0.0077
Children: 2	0.016[a]	0.0038	0.019[a]	0.0044	0.010	0.0074	0.012[a]	0.0057	0.014[a]	0.0068	0.003	0.0104
Children: ≥3	0.018[a]	0.0056	0.025[a]	0.0063	0.000	0.0117	0.002	0.0082	0.011	0.0096	-0.026[c]	0.0159
Constant	0.402[a]	0.0279	0.412[a]	0.0324	0.370[a]	0.0556	0.511[a]	0.0400	0.493[a]	0.0446	0.624[a]	0.0922
FE	NO		NO		NO		YES		YES		YES	
R²	0.23		0.23		0.23		0.19		0.19		0.16	
N	**31.090**		**22.714**		8.158		**31.090**		**22.714**		8.158	

* a) p<0.01, b) p<0.05, c) p<0.1.

Table 4: Decomposition analysis of the health concentrations index, estimated elasticities, concentration indexes, health inequality contributions of regressors in absolute values and % of health scores, and with bootstrapped t-values (based on 1000 replications).[*]

Variables	All complete twin pairs										All complete MZ twin pairs									
	OLS estimation (1)					WTP estimation (2)					OLS estimation (3)					WTP estimation (4)				
	Elast[*]	C	Eff	Eff %	Agg %	Elast[*]	C	Eff	Eff %	Agg %	Elast[*]	C	Eff	Eff %	Agg %	Elast[*]	C	Eff	Eff %	Agg %
C (health)		0.044[a]					0.044[a]					0.043[a]					**0.043**			
Male	0.026[a]	0.268[a]	0.007[a]	15.990	15.990	0.027[a]	0.268[a]	0.007[a]	16.407	16.407	0.025[a]	0.292[a]	0.007[a]	17.017	17.017					
Age 51-60	-0.011[a]	0.108[a]	-0.001[a]	-2.576							-0.009[a]	0.112[a]	-0.001[a]	-2.282						
Age 61-70	-0.010[a]	-0.095[a]	0.001[a]	2.176							-0.009[a]	-0.113[a]	0.001[a]	2.298						
Age 71-84	-0.018[a]	-0.445	0.008[a]	17.649	17.249						-0.016[a]	-0.433[a]	0.007[a]	15.898	15.914					
L(Income)	0.457[a]	0.024[a]	0.011[a]	24.348	24.348	0.321[a]	0.024[a]	0.008[a]	17.121	17.121	0.508[a]	0.023[a]	0.012[a]	27.854	27.854	0.208	0.023[a]	0.005	11.374	11.374
Schooling	0.079[a]	0.059[a]	0.005[a]	10.502	10.502	0.041[a]	0.059[a]	0.002[a]	5.450	5.450	0.068[a]	0.060[a]	0.004[a]	9.599	9.599	0.038	0.060[a]	0.002	5.328	5.328
Working part-time	-0.005[a]	-0.274[a]	0.002[a]	3.390		-0.007[a]	-0.274[a]	0.002[a]	4.617		-0.004[a]	-0.271[a]	0.001[a]	2.747		-0.007[a]	-0.271[a]	0.002[a]	4.296	
Self-employed	0.002[a]	0.022	0.000	0.124		0.002[a]	0.022	0.000	0.095		0.002[a]	-0.016	0.000	-0.083		0.002	-0.016	0.000	-0.091	
Unemployed	-0.001[a]	-0.174[a]	0.000[a]	0.296		-0.001[a]	-0.174[a]	0.000[a]	0.549[a]		-0.001[c]	-0.208[a]	0.000[c]	0.398		-0.002[b]	-0.208[a]	0.000[b]	0.801	
Economically inactive	-0.029[a]	-0.238[a]	0.007[a]	15.655		-0.029[a]	-0.238[a]	0.007[a]	15.808		-0.030[a]	-0.235[a]	0.007[a]	16.256		-0.026[a]	-0.235[a]	0.006[a]	14.237	
Retired	0.000	0.156[a]	0.000	0.045	19.511	-0.001[a]	0.156[a]	0.000[a]	-0.498	21.070	0.000	0.128[a]	0.000	0.041	19.359	-0.002[c]	0.128[a]	0.000[c]	-0.561	18.682
Divorced	-0.004[a]	0.003	0.000	-0.020		-0.002[a]	0.003	0.000	-0.014		-0.004[a]	0.005	0.000	-0.047		-0.002	0.005	0.000	-0.018	
Widowed	-0.003[a]	-0.276[a]	0.001[a]	1.994		-0.001	-0.276[a]	0.000	0.738		-0.003[a]	-0.290[a]	0.001[a]	2.134		0.000	-0.290[a]	0.000	0.124	
Unmarried	-0.004[a]	-0.066[a]	0.000[a]	0.629	2.602	-0.005[a]	-0.066[a]	0.000[a]	0.730	1.454	-0.002[a]	-0.043[a]	0.000[c]	0.246	2.332	-0.004	-0.043[a]	0.000	0.418	0.525
Children: 1	0.001[c]	0.115[a]	0.000[c]	0.199		0.000	0.115[a]	0.000	-0.067		0.000	0.091[a]	0.000	-0.090		-0.003[b]	0.091[a]	0.000[b]	-0.697	
Children: 2	0.001[a]	0.162[a]	0.000[a]	0.542		0.001	0.162[a]	0.000	0.389		0.001	0.163[a]	0.000	0.335		0.000	0.163[a]	0.000	0.101	
Children: ≥3	0.001[a]	0.156[a]	0.000[a]	0.227	0.968	0.000	0.156[a]	0.000	0.019	0.340	0.000	0.150[a]	0.000	0.004	0.249	-0.001	0.150[a]	0.000	-0.330	-0.926
Fixed effects	-	-	-	-		0.656[a]	0.024[a]	0.016[a]	35.361	35.361	-	-	-	-	-	0.798[a]	0.034[a]	0.027[a]	63.997	63.997
Residual			0.004	8.83	8.83			0.001	3.295	3.295			0.004	7.677	7.677			0.001	1.020	1.020
Sum over X's			0.044	100	100			0.044	100	100			0.043	100	100			0.043	100	**100**

Table 5: Representativity of the OLS-based decomposition on the twin sample and ULF sample and a restrictive demographic and socioeconomic model, and with bootstrapped significance test (based on 1000 replications).

Variables	All complete twin pairs					All complete MZ twin pairs					ULF (representative) sample				
	(1)					**(2)**					**(3)**				
	Elast*	C	Eff	Eff %	Agg %	Elast*	C	Eff	Eff %	Agg %	Elast*	C	Eff	Eff %	Agg %
C (health)		0.044 [a]					0.043					0.052[a]			
Male	0.036	0.268	0.010[a]	21.917	21.917	0.035	0.292	0.010[a]	23.610	23.610	0.030	0.188	0.006[a]	11.012	11.012
Age 51-60	-0.022	0.108	-0.002[a]	-5.398		-0.017	0.112	-0.002[a]	-4.495		-0.016	0.190	-0.003[a]	-5.867	
Age 61-70	-0.015	-0.095	0.001[a]	3.319		-0.012	-0.113	0.001[a]	3.195		-0.017	-0.121	0.002[a]	4.105	
Age 71-84	-0.016	-0.445	0.007[a]	15.762	13.683	-0.013	-0.433	0.006[a]	13.394	12.094	-0.031	-0.422	0.013[a]	25.180	23.418
L(Income)	0.626	0.024	0.015[a]	33.392	33.392	0.650	0.023	0.015[a]	35.591	35.591	0.695	0.028	0.020[a]	37.706	37.706
Schooling	0.111	0.059	0.007[a]	14.769	14.769	0.102	0.060	0.006[a]	14.336	14.336	0.104	0.063	0.007[a]	12.715	12.715
Residual			0.007[a]	16.239	16.239			0.006	14.369	14.369			0.008	15.149	15.149
Sum over X's			0.044	0.044	100			0.043	100	100			0.052	100	100

Figure 1: Summary of the percentage contribution of different determinants of income-related inequalities in health.

OLS, all complete twin pairs

4 | 27.85 | 9.599 | 19.359 | 2.332

OLS, all complete MZ twin pairs

WTP, all complete twin pairs

3 | 24.348 | 10.502 | 19.511 | 2.602

WTP, all complete MZ twin pairs

2 | 17.121 | 5.45 | 21.07 | 1.454