Inequalities in self-reported health: validation of a new approach to measurement

by

Eddy van Doorslaer* [a] and Andrew M. Jones [b]

Revised January 2002

[a] Department of Health Policy & Management, Erasmus University, 3000 DR Rotterdam, The Netherlands

[b] Department of Economics and Related Studies, University of York, York YO10 5DD, UK

* Corresponding author. Tel. +31 10 4088555. Fax: +31 10 4089094. E-mail: vandoorslaer@bmg.eur.nl. This paper derives from the project "Economic determinants of the distribution of health and health care in Europe" (known as the ECuity II Project), which is funded in part by the European Community's Biomed II programme (contract BMH4-CT98-3352). We are grateful to the EC for financial support, to Statistics Canada for access to the NPHS data. We are grateful to Mark Dusheiko, Xander Koolman, Adam Wagstaff and participants in the ECuity Workshop in Aix-en-Provence (F), and in seminars at the universities of Bergen, York and McGill in Montreal for useful suggestions and comments on an earlier version of the paper.

Abstract

This paper assesses the internal validity of using the McMaster *Health Utility Index Mark III* (HUI) to scale the responses on the typical self-assessed health (SAH) question "How do you rate your health status in general?". It compares alternative procedures to impose cardinality on the ordinal responses. These include OLS, ordered probit and interval regression approaches. The cardinal measures of health are used to compute and to decompose concentration indices for income-related inequality in health. These results are validated by comparison with the individual variation in the 'benchmark' HUI responses obtained from the Canadian *National Population Health Survey 1994-95*. The interval regression approach, which exploits a mapping from the empirical distribution function of HUI into SAH, outperforms the other approaches. In addition, we show how the method can be extended to allow for differences in SAH thresholds across different groups of people and to measuring and decomposing 'pure' health inequality.

JEL Classification: D30; D31; I10; I12

Keywords: health inequality; health measurement; health-related quality of life; Canada;

1. Introduction

One of the challenges in investigating inequalities in health is that, very often, health information is only available at an ordinal level. One of the most commonly used indicators of overall individual health in general population surveys is the simple question “How is your health in general?”, with response categories ranging from “very good” or “excellent” to “poor” or “very poor”. This categorical variable has been shown to be a very good predictor variable of other outcomes, such as subsequent use of medical care or of mortality (see e.g., Idler and Benyamini, 1997). However it does not provide a cardinal health (utility) scale that can be used, for instance, for quality adjustments of life expectancy. Categorical measures of health create a problem for the measurement of inequalities in health. The health concentration index, and the related slope index of inequality, require information on health in the form of either a continuous variable or a dichotomous variable (Wagstaff and van Doorslaer, 1994).

Although, in health interview surveys, the simple self-assessed health (SAH) question is usually supplemented by a host of other measurement instruments, its use remains very popular in general socioeconomic surveys, such as the European Community Household Panel, which tend to have limited space for health questions. It is often the key health variable available in the largest number of surveys and over the longest period of time. Imposing a cardinal scale on SAH allows concentration indices to be used for measuring and explaining inequalities in health. But this comes at a cost: imposing the wrong scaling may lead to misleading results. The aim of this paper is to compare a range of different methods for scaling SAH against the benchmark provided by a direct measure of health utility.

In the past, researchers concerned with the measurement of inequalities in health have dealt with the ordinal scale problem either by dichotomizing the variable into a healthy/non-healthy distinction or by arbitrarily imposing some sort of scaling assumption. The dichotomization approach has well-known disadvantages. Not all health variation contained in the SAH variable is used and it makes comparisons of inequality over time (e.g. Wagstaff and Van Doorslaer, 1994) or across populations (e.g. Van Doorslaer and Koolman, 2000) unreliable. Results often differ depending on the choice of the cut-off for healthy/non-healthy.

A variety of methods have been adopted in the scaling approach. Wagstaff and Van Doorslaer (1994) assumed that underlying the categorical empirical distribution of the responses to the self-assessed health (SAH) question was a latent, continuous but unobservable health variable with a standard lognormal distribution. This allowed 'scoring' of the SAH categories using the midpoints of the intervals corresponding to the standard lognormal distribution. The advantage of this approach was that it enabled a comparison of results across surveys with differing numbers of response categories to the SAH question (e.g. Van Doorslaer *et al*, 1997). The health inequality results obtained using this scaling procedure also appeared to be comparable to those obtained using truly continuous generic measures like the SF36 (Gerdtham et al, 1999) or the HUI III (Humphries and van Doorslaer, 2000). But an obvious disadvantage is that it imposes the unduly strong assumption of an identical latent health distribution across surveys or countries. Also the latent variable does not line on the [0,1] scale expected of a health utility score.

A second approach is to estimate ordered probit regressions using the SAH categories as the dependent variable and to re-scale the underlying latent variable of this model to compute "quality weights" for health between 0 and 1 (Cutler and Richardson, 1997; Groot, 2000). So far no evidence on the validity of this method has been produced. Finally, some researchers have used external information on the means per SAH category of a more generic health measure from another survey to score the SAH categories in a survey not containing such generic measures (Van Doorslaer and Koolman, 2000; Gerdtham and Johanneson, 2000). An obvious disadvantage of this approach - and indeed the standard lognormal one - is that it uses the SAH information as if it were a continuous variable while in fact the underlying information is still inherently categorical. When regression analysis is used to model the measure of health outcome this leads to the inappropriate use of OLS on what remains essentially a categorical dependent variable and does not exploit the within-category variation in health.

In this paper, we compare the internal validity of some of the latter approaches by exploiting the fact that the Canadian *National Population Health Survey 1994-95* contains both the simple SAH question and one of the well-known measures of health utility. This is the *McMaster Health Utility Index Mark III,* as used in computing healthy life expectancies for Canada (Feeny *et al.,* 1995, Furlong *et al.,* 1998). By assuming that the HUI can serve as an appropriate measure of the *actual* health of individuals, we can assess to what extent the results obtained with the various approaches approximate those obtained with direct observations of HUI.

Our reasoning is as follows. When individual-level data on HUI is available it is possible to use regression analysis on the actual HUI values to carry out analysis of inequality. In particular, regressions can be used to decompose the explained component of HUI using the methods of Wagstaff, van Doorslaer and Watanabe (2002). Now imagine that the survey includes SAH but not HUI. If individual level data on HUI is not available it would not possible to link a particular individual's SAH and HUI. However, imagine we do have external information on the shape of the empirical distribution function (EDF) of HUI in the relevant population (say from another survey). Our aim is to use this information and find a latent variable method that approximates the predictions from a regression on actual HUI. This gives us a criterion for assessing the validity of the methods and means that our benchmark is to see how the predictions from different methods compare to the predictions from a regression on actual HUI.

We also explore the importance of the phenomenon which has been termed 'state-dependent reporting bias' (Kerkhofs and Lindeboom, 1995), 'scale of reference bias' (Groot, 2000) and 'response category cut-point shift' (Sadana *et al*, 2000, Murray *et al*, 2001) for the measurement of income-related health inequality. This occurs if sub-groups of the population use systematically different threshold levels for SAH, despite having the same level of 'true' health. These differences may be influenced by, among other things, age, sex, education, language and personal experience of illness. We test for the existence of this effect and we illustrate the impact of allowing for the thresholds in the interval regression approach to depend on disability.

Finally, it has been argued (e.g. Gakidou *et al*, 2000) that *all* health inequalities can to some extent be a cause of concern, not just those which display a systematic relationship with indicators of socio-economic status. Systematic health disparities have been shown to exist not only with respect to variables like income and education, but also with respect to place of residence, race, marital status, ethnic origin and a host of other characteristics of groups or individuals which health policy makers may find relevant. Consequently, it would be of interest to be able to compute measures of total or 'pure' inequality in health and decompose them into their sources, including socioeconomic factors like income. We will illustrate how the methods proposed in this paper also allow for such decompositions.

2. Methods

2.1 Measurement of health

Econometric analysis of an ordered categorical dependent variable, such as SAH, is typically based on the ordered probit or logit model or, if information on the scaling of the variable is available, the interval (grouped data) regression model.

2.1.1 The ordered probit model

The ordered probit model can be used to model a discrete dependent variable that takes ordered multinomial outcomes for each individual *i*, for example $y_i = 1,2......,m$. This applies to our measure of self-assessed health (SAH), which has categorical outcomes poor, fair, good, very good and excellent. The model can be expressed as,

$$(1) \qquad y_i = j \quad if \;\; \mu_{j-1} < y^*_i \leq \mu_j \;\;, \quad j = 1,......,m$$

where the latent variable, y^*, is assumed to be a function of a vector of socio-economic variables x,

$$(2) \qquad y^*_i = x_i\beta + \varepsilon_i \;\;, \qquad \varepsilon_i \sim N(0,1)$$

and $\mu_0 = -\infty$, $\mu_j \leq \mu_{j+1}$, $\mu_m = \infty$. Given the assumption that the error term is normally distributed, the probability of observing a particular value of y is,

$$(3) \qquad P_{ij} = P(y_i = j) = \Phi(\mu_j - x_i\beta) - \Phi(\mu_{j-1} - x_i\beta)$$

where $\Phi(.)$ is the standard normal distribution function. With independent observations, the log-likelihood for the ordered probit model takes the form,

$$(4) \qquad LogL = \sum_i \sum_j y_{ij} \log P_{ij}$$

where the y_{ij} are binary variables that equal 1 if $y_i = j$. This can be maximised to give estimates of β and of the unknown threshold values μ_j.

Predictions of the linear index, $x_i\beta$, can then be used as a measure of individual health and, after appropriate re-scaling, be used as ‘quality weights’ or utility proxies. Cutler and Richardson (1997) and Groot (2000) have proposed to do this by using the re-scaled linear index $y^*/(\mu_{m-1}-\mu_1)$. With their normalisation that $\mu_1=0$, this implies that their “QALY score” equals 0 when $y^*=\mu_1$, when the latent index is equal to the lowest threshold, for example between poor and fair health. It equals 1 when $y^*=\mu_{m-1}$, at the highest threshold, for example between very good and excellent health. A drawback of this re-scaling procedure is that individuals with predicted values of the latent index below the bottom threshold should receive negative quality-of-life or utility weights, while individuals with values above the top threshold should receive weights greater than one. This seems undesirable and can easily be remedied by adopting an alternative scale. We have considered the following two alternatives:

i. The predictions from the ordered probit model can be re-scaled to the [0,1] interval. Let y^1 be the predicted linear index from the ordered probit model and let y^{max} be the largest individual prediction and y^{min} the smallest. Then the re-scaled variable can be calculated as,

$$y^2 = (y^1 - y^{min})/(y^{max} - y^{min}) \tag{5}$$

ii. An alternative is to re-scale to *the actual observable range* of HUI scores in the NPHS. In this case, using the observed range (0.031-1) would make little difference. But it is possible to use the range of average HUI values for the demographic groups defined by the set of regressors, *x*. One way to do this is to use the minimum and maximum predictions from an OLS regression of HUI on *X* (say, p^{min} and p^{max}),

$$y^3 = p^{min} + (p^{max} - p^{min})y^2 \tag{6}$$

This re-scales from the interval [0,1] to the interval $[p^{min}, p^{max}]$. It can be interpreted as the observable range of HUI, conditional on the set of regressors, *x*. In practice, of course, this is only practicable for samples where a measure like HUI is available.

2.1.2 The interval regression model

Interval, or grouped data, regression provides an alternative to the ordered probit model in the case when the values of the upper and lower limits of the intervals are known. Because the m's are known, the estimates of b are more efficient and it is possible to identify the variance of the error term s^2 and, hence, the scale of y^* (see e.g., Jones, 2000). Software to estimate this model is readily available, for example in Stata and in Limdep (where it is called grouped data regression).

Our approach is to use HUI scores to scale the intervals of SAH. To do this we assume that there is a stable mapping from HUI to the (latent) variable that determines reported SAH and that this applies for all individuals. This implies that an individual's rank according to HUI will correspond to their rank according to SAH and, hence, the q-th quantile of the distribution of HUI will correspond to the q-th quantile of the distribution of SAH. We adopt a nonparametric approach to estimate the thresholds (m_j). The first step is to compute the cumulative frequency of observations for each category of SAH. Then find the quantiles of the empirical distribution function (EDF) for HUI that match these frequencies. More formally,

$$m_j = F^{-1}(G_j) \tag{7}$$

where $F^{-1}(.)$ is the inverse of the EDF of HUI and G_j is the cumulative frequency of observations for category j of SAH.

Because we use HUI values to scale the thresholds for SAH, the linear index for the interval regression model is measured on the same scale. The unconditional prediction of the linear index $x_i b$ gives us a prediction of each individual's level of HUI derived from their observed SAH. This is the level of HUI that would be predicted knowing that an individual has characteristics x. The prediction is both continuous and linear in the x's. Linearity is a useful property that means that concentration indices calculated using the predictions are suitable for decomposition analysis (see section 2.3 below).

An alternative way of computing the predicted values from the interval regression model is to use the expected value of the linear index, conditional on the individual's observed category of SAH,

(8) $E(y^*_i \mid x_i, \mu_{j-1} < y^*_i \leq \mu_j) =$

$$x_i\beta + \sigma\left\{\phi((\mu_{j-1}-x_i\beta)/\sigma) - \phi((\mu_j-x_i\beta)/\sigma)\right\}/\left\{\Phi((\mu_j-x_i\beta)/\sigma) - \Phi((\mu_{j-1}-x_i\beta)/\sigma)\right\}$$

This gives the level of HUI that would be predicted knowing both x *and* the category of SAH that the individual reports. Comparing these conditional predictions to the actual data on HUI is a useful way of assessing the reliability of the interval regression method.

A priori, the interval regression approach appears to have several advantages over alternative methods. First, using either interval regression or an ordered probit model means that decomposition analysis does not have to be based on the inappropriate use of OLS to model a categorical dependent variable. Second, using the predicted linear index, rather than simply using the mean HUI score for each category of SAH, leads to greater individual-level variation in the measure of health. This is especially true if x includes continuous variables such as income. In effect the between-SAH category variation is exploited to generate some within-SAH category variation in HUI, although HUI itself is unobserved. Third, interval regression, like the category means method but unlike the ordered probit model, allows for the incorporation of external information to scale the categorical observations of SAH. As such, the predictions are measured on the same scale as HUI and do not require ex post re-scaling, as is often done with ordered probit predictions. Finally, the thresholds used in the interval regressions can be allowed to be different for different groups of individuals or when comparing across different countries (depending on the relative frequencies in each category of SAH). As the thresholds determine the scale of the latent variable, this is equivalent to allowing for heteroskedasticity in the latent variable specification.

2.1.3 Testing the internal validity of the interval regression approach

In order to assess how the ordered probit and interval regression methods perform in terms of their internal validity, it is possible to exploit the fact that we have the actual HUI values to use as a benchmark in this study. Our strategy is to compare the descriptive performance (measured by the mean, standard deviation, 25th, 50th and 75th percentiles, minimum and maximum) and the measured degree of inequality (measured by the concentration and Gini indices) of the various measures. We also perform a decomposition analysis of each of these alternative health indices. This allows us to assess whether the different health measures have an impact on the measurement and explanation of inequalities in health. The first two are based on the actual HUI data:

(i) the *actual* measured HUI scores per individual,

(ii) the OLS predicted HUI scores based on the actual observations,

The others use indirect methods:

(iii) the *attributed* mean HUI score for each category of SAH.

(iv) the re-scaled predictions of the linear index from an ordered probit model,

(v) the unconditional predictions of the linear index from an interval regression model.

Method (ii) provides the benchmark for assessing the performance of the predictions in methods (iii)-(v).

It is sometimes argued that the mapping of 'true health' into SAH categories varies with respondent characteristics. This has been referred to as 'state-dependent reporting bias' (Kerkhofs and Lindeboom, 1995) or 'scale of reference bias' (Groot, 2000) and can be formally tested by making the threshold levels dependent on some or all of the exogenous variables used in the ordered probit model. It is worth noting that allowing the scaling of SAH to vary across individual characteristics is equivalent to a heteroskedastic specification. To illustrate, let,

$$(9) \qquad y^*_i = f(x_i) + e_i \, , \qquad e_i \sim N(0, s g(x_i))$$

for some general functions $f(.)$ and $g(.)$. Then, adapting (3),

$$(10) \qquad P_{ij} = P(y_i = j) = F[\{m_j - f(x_i)\}/s g(x_i)] - F[\{m_{j-1} - f(x_i)\}/s g(x_i)]$$

$$= F[\{m_j/g(x_i) - f(x_i))/g(x_i)\}/s] - F[\{m_{j-1}/g(x_i) - f(x_i))/g(x_i)\}/s]$$

$$= F[\{m_{i,j} - h(x_i)\}/s] - F[\{m_{i,j-1} - h(x_i)\}/s]$$

which gives an interval regression with thresholds that vary across individual characteristics, x_i.

As mentioned above, a limitation of the ordered probit model is that it does not allow the incorporation of external information on the HUI distribution. This is a limitation for our analysis, which is concerned with whether state-dependent reporting bias exists *after* allowing for an individual's level of HUI. Nevertheless a comparison of the standard ordered probit model, as in (3), and one in which the bs are allowed to vary across categories (b_j) provides some preliminary evidence.

Our real concern is whether the mapping from HUI to SAH is stable across different groups of individuals. With our nonparametric approach it is possible to compute the empirical distribution function of HUI for different sub-samples and see whether the thresholds corresponding to the cumulative frequencies for SAH are stable across groups. We compare the HUI means and interval limit values for different age-sex categories and for those with and without disabilities. Of course, this does not guarantee that mapping will be stable across other sub-groups or across different populations. However, we show how the interval regression method can be extended to allow for individual- or group-specific thresholds, if there is evidence that the thresholds vary.

2.2. Measurement of inequality

We use the *health concentration index* (Wagstaff, van Doorslaer and Paci, 1989) as our measure of *relative* income-related health inequality. Suppose we have a continuous cardinal measure of health (utility) y_i. The concentration curve $L(s)$ plots the cumulative proportion of the population (ranked by income, beginning with the lowest incomes) against the cumulative proportion of health. If $L(s)$ coincides with the diagonal, everyone enjoys the same health. If, by contrast, $L(s)$ lies *below* the diagonal, inequalities in health exist and favour the richer members of society. The further $L(s)$ lies from the diagonal, the greater the degree of inequality. The health concentration index, C, is defined as twice the area between $L(s)$ and the diagonal. C takes a value of zero when $L(s)$ coincides with the diagonal and is negative (positive) when $L(s)$ lies above (below) the diagonal. The minimum and maximum values of C using individual-level data are -1 and +1 respectively: these occur when all the population's ill-health is concentrated in the hands of the most and least disadvantaged persons respectively.

For weighted data, the computation formula for C given by Kakwani, Wagstaff and van Doorslaer (1994) can be modified as follows:

$$C = \frac{2}{m} \sum_{i=1}^{N} w_i y_i R_i - 1 \qquad (11)$$

where,

$$m = \sum_{i=1}^{N} w_i y_i \qquad (12)$$

is the (weighted) mean health of the sample, N is the sample size, w_i is the sampling weight of individual i (with the sum of w_i equal to N), and R_i is the (weighted) relative fractional rank of the ith individual. The latter is defined as (Lerman and Yitzhaki, 1989)

$$R_i = \tfrac{1}{N} \sum_{j=1}^{i-1} w_j + \tfrac{1}{2} w_i \quad \text{where } w_0 = 0 \qquad (13)$$

and thus indicates the weighted cumulative proportion of the population up to the midpoint of each individual weight. As Kakwani *et al.* (1994) show, C can alternatively be derived as the estimate of γ in the following convenient WLS regression:

(14) $$2s_R^2 \left[y_i / m \right] \sqrt{w_i} = a \sqrt{w_i} + g\, R_i \sqrt{w_i} + u_i \,.$$

where $s_R^2 = \frac{1}{N} \sum_{i=1}^{N} w_i \left(R_i - \frac{1}{2} \right)^2$ is the (weighted) variance of R_i. The estimator of γ is equal to,

(15) $$\hat{g} = \frac{2}{m} \sum_{i=1}^{N} w_i \left(y_i - m \right) \left(R_i - \tfrac{1}{2} \right),$$

which, given that the mean of R_i is equal to ½, gives the result $\hat{g} = C$.

This means that C can be computed conveniently using the weighted covariance of m and the (weighted) fractional rank (Lerman and Yitzhaki, 1989) as:

(16) $$C = \frac{2}{m} \sum_{i=1}^{N} w_i \left(y_i - m \right) \left(R_i - \tfrac{1}{2} \right) = \frac{2}{m} \; w\operatorname{cov}(y_i, R_i)$$

where *wcov* denotes the weighted covariance.

An advantage of using the predictions from ordered probits or interval regressions is that they can then be used to look at *total* health inequality, including inequality *not* specifically linked to income. This allows for the broader analysis of health pseudo-Lorenz curves and the measurement of inequality using a *Gini coefficient* of health inequality G (e.g. Le Grand, 1989, and Wagstaff, Paci and van Doorslaer, 1991). It can be defined analogously to the health concentration index as,

(17) $$G = \frac{2}{m} \sum_{i=1}^{N} w_i y_i R'_i - 1$$

where all other variables are as before but R'_i now denotes the relative (weighted) fractional rank in the *health* distribution with individuals ranked from lowest to highest health.

Standard errors for C and G can be obtained from estimating equations like (14) by least squares methods. A more accurate estimator for the standard error of C, taking into account the serial correlation in the errors and the dependence of the observations as a result of the presence of the relative rank variable on the right-hand side of (14), has been developed by Kakwani *et al*. (1994). But since this estimator does not correct for potential heteroskedasticity and for the fact that the data were taken from a clustered sampling design, we have chosen not to use those here but rather to make use of the Huber-White robust estimator of the variance matrix.

2.3 Decomposing inequality

Measuring inequality is useful and important, but more interesting for policy purposes is to unravel and quantify the contributions of various determinants of health to measured inequality. One straightforward way of doing this is demonstrated by Wagstaff, van Doorslaer and Watanabe (2002). They show that, for a linear regression model,

$$y_i = a + \sum_k b_k x_{ki} + e_i \,, \tag{18}$$

the concentration index for y can be written as,

$$C = \sum_k (b_k \bar{x}_k / m) C_k + GC_e / m = C_{\hat{y}} + GC_e / m \,, \tag{19}$$

where m is the mean of y, $\bar{x}_k$ is the mean of x_k, C_k is the concentration index for x_k and GC_e is the generalized concentration index for e_i.

Equation (19) shows that C can be thought of as being made up of two components. The first is the deterministic, or “explained”, component. This is equal to a weighted sum of the concentration indices of the regressors, where the weights are simply the elasticities of y with respect to each x_k. The second is a residual, or “unexplained”, component. This reflects the inequality in health that cannot be explained by systematic variation in the x_k across income groups. The decomposition shows how each determinant’s separate contribution to explained income-related health inequality can be decomposed (i) its health elasticity ($b_k \bar{x}_k / m$) and (ii) its income-related inequality (C_k). This allows us further decompose each factor's contribution into these two terms

Equation (19) can also be used to decompose inequality as measured by the Gini in exactly the same way. This can be computed directly, by replacing income rank by health rank. Alternatively, it is possible to make use of the relationship between the Gini coefficient and the concentration index (Kakwani, 1980, p.174),

$$G = \frac{\rho(y, R'_i)}{\rho(y, R_i)} C \tag{20}$$

where ρ(.) denotes a correlation coefficient. Substituting for C from (19) gives the decomposition for the Gini. We will exploit this useful property in the analysis of inequality in the self-reported health of Canadians in section 4.

The decomposition result in (19) relies on the fact that *y* is additive in its components *x*. Because the predicted linear indexes from the ordered probit or the interval regression model are also additive in the *x*'s, the same kind of decomposition analysis can be applied to them. This is a further attraction of using the latent variable approach to deal with the categorical measure of SAH. This kind of decomposition analysis could not be applied directly to the observed categorical measure of health. However, it is worth pointing out that the use of the predicted linear index means that it is only the *explained* variation in the health measure that can decomposed.

3. Data and variable definitions

The data used in this paper are taken from the first wave (in 1994-1995) of the Canadian *National Population Health Survey (NPHS)*. The target population of the *NPHS* includes household residents in all provinces, with the exclusion of populations on Indian Reserves, Canadian Forces Bases and some remote areas of Ontario and Quebec. A total of 26,430 households were selected for the survey. In each household, a randomly selected household member, aged 12 years or older, was selected for a more in-depth interview. This interview included questions on health status, risk factors, and demographic and socio-economic information. The data were weighted using the survey weights to adjust for the complex multi-cluster sample design of the NPHS. Detailed information about the *NPHS* content and sample design has been published elsewhere (e.g. Tambay and Catlin, 1995).

The two key variables for this study are self-assessed health (SAH) and health status as measured by the Health Utility Index (HUI). As part of the in-depth component of the *NPHS*, respondents were asked: " In general, how would you say your health is?" The response categories were excellent, very good, good, fair, and poor. Also, each respondent was assigned a Health Utility Index score based on their response to the questions of the eight-attribute Health Utility Index Mark III health status classification system. The Health Utility Index is a generic health status index, developed at McMaster University that measures both quantitative and qualitative aspects of health (Torrance *et al*, 1995, 1996; Feeny *et al*, 1995). It provides a description of an individual's overall functional health, based on eight attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. The Health Utility Index assigns a single numerical value, between zero and one, for all possible combinations of levels of these eight self-reported health attributes. A score of one indicates perfect health. The Health Utility Index also embodies the views of society concerning health status, inasmuch as preferences about various health states are elicited from a representative sample of individuals.

The key ranking variable used is total income before taxes and deductions, as measured in the NPHS as a categorical variable with 11 response categories. For the purposes of this study, the two lowest income groups- no income and less than $5,000- were combined into one group, thus reducing the number of income categories from 11 to 10. The midpoint of each income category was then attributed to all households in that category and subsequently divided by an equivalence factor equal to (number of household members)$^{0.5}$, to adjust for differences in household size. The income values assigned for the top and bottom groups were Can $2,500 and $87,500.00 respectively.

Other health determinants included in the analysis are the following. (i) Education level, the highest level of general or higher education completed is available at three levels: recognised third level education (ISCED 5-7), second stage of secondary level of education (ISCED 3) and less than second stage of secondary education (ISCED 0-2)); (ii) Marital status distinguishes between married, separated/divorced, widowed and unmarried (including co-habiting); (iii) Activity status includes employed, self-employed, student, unemployed, retired, housework and 'other economically inactive'.

The NPHS has a complex multi-stage stratified sampling design. In order to keep the sample representative of the Canadian adult population, sampling weights are used in all analyses.

4. Results

4.1 The distributions of SAH and HUI

It is well known that the health of a general population sample has a very skewed distribution, with the great majority of respondents reporting their health to be good to excellent. Table 1 provides the NPHS (weighted) distributions of SAH and the associated HUI scores. It illustrates that only 11.5% of Canadian adults report lower-than-good self-assessed health. The skewness is also illustrated by the fact that the 'distances' between excellent health, very good health and good health in terms of HUI means are much smaller than between poor and fair health. In fact, the loss of health utility increases with each step in SAH when going down from excellent to poor health. This suggests that the health differences between SAH categories increase with lowering SAH categories. Also the standard deviation of the mean HUI increases with lowering levels of SAH. By collapsing health measurement into just 5 categories, these two differences are no longer visible.

By applying the mapping procedure described in section 2 and by using the cumulative distribution function of HUI as the benchmark, we can derive the thresholds which define the HUI intervals corresponding to each SAH level. This is also illustrated in Figure 1 which shows how the actually observed frequencies of each SAH level are used to determine the corresponding interval boundaries in HUI units. Note that this procedure assumes that, for instance, all individuals reporting poor health have HUI scores below 0.428. We can see in column 5 of Table 1 that the assumption is violated because in reality the observed mean HUI for this group equal to 0.557 is above this interval threshold.

Our method makes the "instrumental" assumption that there is a stable monotonically increasing mapping from HUI to (latent) SAH. This assumption is instrumental in the sense that it provides us with a way of generating values for the thresholds that are used in the interval regression. The validity of this approach should be judged by the performance of the interval regression method against the benchmark provided by predictions from regressions on actual HUI data. There are problems with trying to test this assumption directly. Unlike HUI, latent SAH is not observed. The actual categorical values of SAH can be used (allowing only five possible ranks). Alternatively, to proxy the conditional expectation of the latent variable, we can use the fitted values from an ordered probit (which, unlike the interval regression, are not tied to HUI in any way). In reality there is likely to be "noise" in the relationship, due to heterogeneity and measurement error, so that it cannot be expected to hold exactly for all individuals. If the relationship is expected to hold on average we can compare the rank of the predicted values from OLS on actual HUI to the rank of the predicted values from the ordered probit. The Spearman rank correlation between actual HUI and actual SAH is 0.4365 and independence is rejected (p=0.0000). The Spearman rank correlation between predicted HUI and predicted y* from the ordered probit is 0.9493. For completeness, it is worth noting that the Spearman rank correlation between predicted HUI and the predictions from the interval regression is 0.9592.

Table 2 presents the same information for four demographic groups: younger men (between 18 and 44), older men (over 45), and younger and older women. It can be seen that there are some differences in mean HUI value by level of SAH, with women often reporting slightly lower values than men, and older respondents reporting somewhat lower HUI values than younger. Obviously, the distribution of SAH is also more concentrated to the right for the (healthier) younger groups. The relevant interval boundaries, however, do not differ greatly across demographic groups, indicating that the interval regression approach is unlikely to be sensitive to making the interval boundaries age-sex specific.

Table 2 also presents the same information on HUI by SAH for two very different sub-populations: those which do and do which do not report to be “on disability or recovering from illness”. Since the ‘disabled’ group includes most people unable to work due to disability (or illness), they may have somewhat divergent views of their self-reported general health conditional on HUI due to coping mechanisms or other forms of adaptation to their disability. We can observe that, on average, and as expected, the disabled group is much more concentrated in the lower SAH groups and the corresponding HUI mean values reported by level of SAH are much lower. Nevertheless, the interval boundaries deriving from the mapping of SAH into HUI do not differ dramatically between the two groups.

To further explore the question of state-dependent reporting bias we compared the standard ordered probit model with one in which all of the βs are allowed to vary across categories of SAH. This suggests that heterogeneity is – in general - not a problem within the Canadian sample: the likelihood ratio test for the null that the slope coefficients are equal across categories is 2[-19,757.979-(-19,789.982)] = 64 (p=0.999), suggesting that this hypothesis cannot be rejected. In what follows, we have therefore ignored the issue of state-dependent reporting bias because interval regressions results with state-dependent interval boundaries gave almost identical results to those with uniform interval boundaries across the sample.

4.2 Comparison of different health measures

In this section we compare the descriptive performance of the regression-based HUI measures with those for the actually observed HUI measure. The intention is to examine to what extent the HUI observations can be approximated by the predicted values of the regression-based approaches. Table 3a presents means, standard deviations, minimum and maximum for the total sample predictions. The actual measured HUI is then compared to four types of predictions obtained from:

(i) an OLS regression using actual HUI as the dependent variable – our benchmark,

(ii) an OLS regression on mean HUI per category of SAH,

(iii) an ordered probit regression using SAH as dependent variable,

(iv) an interval regression using SAH as dependent variable and HUI-based threshold levels.

The ordered probit predictions from (iii) have been re-scaled in order to be comparable on the HUI scale in terms of health utility units.

We can observe that much of the inter-individual variability in health (HUI) is lost when going from the actual to (any type of) predicted values, even on the basis of a regression on actually observed HUI scores. But of the three possible candidates, the interval regression approach comes closest to the descriptive statistics obtained from the benchmark provided by the OLS regression on actual HUI scores. The predicted mean is not identical to the observed mean, but the variability is higher and the predicted range and percentiles are closer to the predicted actual ones than in any of the other options. The application of OLS to HUI-mean-scored SAH categories shows less variability and predicts over a narrower range. The ordered probit model, even after using the simple re-scaling to the [0,1] interval, does not approximate the linear predictions very well at all. The more appropriate re-scaling to $[\pi^{min},\pi^{max}]$ gets us somewhat closer, but is not usually an option given that this interval cannot be observed in the absence of HUI scores. In sum, the interval regression approach seems to provide the closest possible approximation to the predictions that would be obtained from an OLS regression on actual HUI scores.

The same descriptive statistics are presented in Table 3b for the *conditional* predictions given the observed level of SAH. Again we observe that the conditional predictions obtained from the interval regression are closest to the linear predictions of the actual HUI score. The mean, standard deviation, range and percentiles of the interval regression predictions are all closer to the ones obtained using OLS on the actual HUI observations than in any of the alternatives, thereby confirming the superior predictive performance of the interval regression approach.

4.3 Inequality and decomposition analysis

4.3.1 Income-related health inequality

The last column of Table 3(a) presents the estimated health concentration indices computed using equation (14). The concentration index for the actual HUI data equals 0.0141. When the actual data is replaced with within-category means (as in van Doorslaer and Koolman, 2000, and Gerdtham and Johanneson, 2000) the estimated concentration index is only 0.0097, showing how the neglect of within-category variation leads to a lower level of measured inequality. The estimates of the overall concentration index can be compared to estimates of the concentration index for the explained variation in health. Our benchmark, the OLS predicted value based on actual HUI, is 0.0135. Using the decomposition in equation (19), this implies that the unexplained component of income-related inequality is small (0.0141-0.0135=0.0006). The interval regression prediction is somewhat higher, at 0.0151, but is closer than any of the estimates generated by the other approximation methods. The estimate based on the OLS predictions using HUI means per SAH category (0.0090) underestimates the explained inequality by around 33%. It is clear that the ordered probit regression does not allow for any sensible approximation of the true degree of inequality, not even after re-scaling on the basis of the interval extremes, which are not usually available.

Essentially the same observations are illustrated by the cumulative health distributions presented in Figure 2. It shows the health concentration curves for each of the HUI-based measures but expressed as the percentage deviations from the diagonal in order to amplify the differences. They are all below the diagonal, indicating inequality favouring the better-off, with the area between the curve and the diagonal serving as a measure of the degree of inequality. But the implied health distributions are not identical. We see that the concentration curves of the OLS and the interval regression predictions are much closer to that of the actual HUI distribution, while the predictions of the means-based OLS leads to a substantial under-estimation and the re-scaled ordered probit to an over-estimation of income-related health inequality.

The regression results used for the decomposition analysis are presented in Table 4. The first two columns present the estimated means and concentration indices for each of the explanatory variables used in the regression equation. The regression results are complemented by a column of inequality contributions expressed as a percentage of $C_{\hat{y}}$. Each term can be interpreted as a sort of 'attributable fraction' of inequality in the predicted $\hat{y}$.

Some caution is required in giving a causal interpretation to the results of the regression analysis. The results tell us about the association between health and factors such as income and activity status. It is, of course, possible that this association reflects reverse causality, namely that good health has a positive effect on income or that both are jointly determined by another third factor. Our estimates are derived from cross-section observational data and issues of causality are better explored with longitudinal or experimental data. However the decomposition methods presented here would be applicable with such data. Our results are intended to illustrate the decomposition methods and should be interpreted as a decomposition of the association between income and health.

An important requirement for any explanatory variable to have an impact on the distribution of health across income is that it is itself unequally distributed across income levels. As such, the negative concentration indices illustrate clearly that groups which are disproportionately represented among the lower income groups in Canada include the lower educated, the unemployed, the disabled, the retired, the divorced and the elderly. The concentration index of the logarithm of income can be seen as a measure of the inequality in the Canadian income distribution. In terms of regression coefficients, we see that the results are remarkably similar across the various approaches. Income invariably shows a statistically significant positive partial effect on health, even after controlling for all these other regressors. Higher income shows a positive but decreasing effect on better health. Other important correlates of adult health are age above 40, education, and activity status.

The magnitude of each variable's contribution to inequality depends on the magnitude of each of its components. Not surprisingly, the unequal distribution of income itself is the largest single contributor to explained health inequality by income. It accounts for 30% of the inequality in the actual OLS-predicted HUI, and for over 40% of the inequality in any of the other predicted HUI measures. It is not unlikely that the fact that income is the only continuous explanatory variable in these regressions contributes to the apparent overestimation of its contribution. The variable with the second largest negative association with health is activity status. Disability status, for instance, is one of the variables with the most negative association with health and the most negative concentration index, highlighting the unfavourable health and income position of disabled individuals. Despite the fact that it affects only 2.8% of the Canadian adult population, it is nevertheless the second most important contributor to health inequality, after income itself. Other variables with important (negative) partial contributions include low education and retirement.

It is striking that the percentage contributions of the interval regression and the OLS- category-means predicted health inequality are very similar, despite the fact that the latter method underestimates health inequality. In fact, using the decomposition method in (19), it is straightforward to show that the percentage contributions to a concentration index are invariant to linear transformations of the index. This is a corollary of the result that, for a variable y with mean μ and concentration index C, any linear transformation $y^*=a + by$ has a concentration index,

(21) $$C^* = \frac{bm}{a+bm} C$$

(see e.g., Kakwani, 1980, p. 176). Combining (21) with the decomposition result in (19) shows that the percentage contribution of a component, *x*, to C will equal its percentage contribution to C*. (21) also demonstrates the invariance of the concentration index to a proportionate re-scaling of health, that is, when *a* equals zero, C* equals C. (21) helps to explain our results. We ran simple linear regressions on the predicted values to gauge whether our different measures of health are close to linear transformations. For the regression of the predicted values for the interval regression on the predictions from the ordered probit model the R^2 is 0.94; from the regression on the predictions for the HUI category means it is 0.998; and for the regression on the actual HUI predictions it is 0.954. This shows that predicted indices are highly collinear with each other and explains why the estimated percentage contributions are very similar. All approximation methods appear to overestimate the percentage contribution of the household's income and the individual's education, at the expense of other determinants' contributions, but this problem is more pronounced for the ordered probit based estimates.

4.3.2. Total health inequality

As outlined in section 2, the regression-based individual health predictions can also be used to examine total inequality in health, irrespective of income or socio-economic position, by ranking all individuals by their (predicted) health and computing Gini indices for health. This would not be possible with the observed categorical measure of SAH but is possible using the predictions from the ordered probit or interval regression models, which introduce a sufficient degree of individual-level variation in the measure of health. We know from equation (20) that health Ginis will always be higher than health concentration indices, as long as the health and income rankings differ (cf eq. (20)). Because income is one of the explanatory variables, our regression models explain more of the income-related variation in health than of the total variation in health.

This is illustrated in Table 5. The top row shows, first of all, that the Gini of actual HUI observations equals 0.0675, which is much higher than the corresponding concentration index in Table 4, which was only 0.0141. It is also twice as high as the Gini index measuring inequality in (OLS) predicted actual HUI scores, which is 0.0326. The difference between these two estimates is attributable to determinants of health that are not captured by the model and to random inter-individual variation. Again, we observe that – of the three proxy predictions used – the interval regression comes closest in approximating the explained inequality in health based on the actual HUI (0.0326). Its Gini is 0.0309, whereas the OLS on the categorical HUI means results in a value of only 0.0185, and the re-scaled ordered probit estimates produce a Gini of 0.0461. All predicted Ginis are only twice the size of the predicted concentration indices in Table 4. This reflects the fact that income-related health inequality accounts for a larger share of predictable total health inequality than of actual inequality.

Ginis for total inequality in predicted health can be decomposed in the same way as we did for concentration indices in Table 4. The regression coefficients now need to be combined with the concentration index of each determinant in terms of health. These are to be interpreted as a sort of 'inverse' concentration index, measuring the association of the regressor with health rank. A positive concentration index indicates that the variable in question is positively associated with health, whereas the reverse holds for a negative concentration index. The ranking is now dependent on predicted health and therefore (slightly) different in each case. The concentration indices in terms of *predicted* health for all health determinants are always higher (in absolute value) than those in terms of *actual* health. This simply reflects the fact that health is predicted using precisely these determinants. As such, they have the association with health rank built in. Regressors with particularly high concentration indices in terms of actual health are, again, variables like disability and very old age. Note also that the concentration index of income with respect to health rank is to be interpreted as a sort of 'mirror image' of the concentration indices of health with respect to income rank used in the previous sections.

The partial contribution of health-related inequality in income as measured by the 'inverse' concentration index is only about 5% of total health inequality as predicted by an OLS regression on actual HUI. This percentage is (more than) doubled if any of the other approaches for predicting HUI is used. Overestimation of the relative contribution (in percentage terms) also occurs for education, at the expense of the relative contribution of women's age, which appears underestimated. In general, and despite the differences in the overall magnitude of inequality generated by the various prediction methods, the decompositions in relative contributions are very similar. This reflects the minor differences in the influence of the regressors across methods. The superiority of the interval regression approach in terms of predicted inequality, therefore, does not carry over to the decomposition analysis.

5. Summary and conclusions

This paper assessed the internal validity of using the McMaster *Health Utility Index Mark III* (HUI) to scale the responses on the typical self-assessed health (SAH) question "How do you rate your health status in general?". It compared alternative procedures to impose cardinality on the ordinal responses obtained. These included OLS, ordered probit and interval regression approaches. Inequality and decomposition results were validated by comparison with the 'benchmark' HUI responses obtained in the Canadian *National Population Health Survey 1994-95*. A note of caution here is that HUI in itself may underestimate the true variability in health status, since there will be additional variability within each HUI category and heterogeneity in the valuation of health states. This may be offset by random measurement error in the HUI classification, which would lead to an overestimate of the true variability.

It is found that an interval regression approach, which assumes and exploits a monotonic mapping of the empirical distribution function from HUI into SAH, performs better than OLS and ordered probit approaches in terms of internal validity. It outperforms the other approaches in the sense that the conditional and unconditional descriptive statistics are closer to those for predictions based on the actual HUI data and that the magnitude of the concentration index computed using this method is closer to the index calculated using the actual HUI scores. The ordered probit overestimates the true degree of health inequality while the attribution of mean generic scores to SAH categories leads to underestimation.

We showed how the method can be extended to allow for differences in SAH thresholds across sub-groups of individuals and to measuring and decomposing 'pure' health inequality. Also the estimates of the Gini coefficient for health are closest to that for actual HUI when predictions from the interval regression are used. The comparative advantage of the interval regression approach does not carry over to the decomposition analysis. The estimates of the relative contributions to income-related and 'pure' inequality in health are very similar across methods. This is because the relative contributions are invariant to linear transformations and, in our data, the different measures of health are highly collinear. All regression-based methods overestimate the contribution of income to both income-related and 'pure' inequalities in health when compared to the actual data.

The application to the Canadian data for 1994 confirms the earlier finding (e.g. Humphries and van Doorslaer, 2000) that significant inequalities in self-reported health by income exist and favour the rich. But it also provides some evidence on the factors which contribute to this finding. For any potential determinant to have an important contribution to the existence of income-related health inequalities, it is required that (a) this factor has a substantial partial health elasticity and (b) that it is unequally distributed across the income distribution. In particular, the analysis showed that around 30-40% of these inequalities is due to the fact that income itself is unequally distributed and has an independent and significant association with health status. Among the other variables included in the health equation, non-working status (especially disability and retirement) and educational status emerge as important other contributing factors. Because of the emphasis on the health measurement issues, no further attempts were undertaken in this paper to test and account for the potential endogeneity of some of these explanatory variables.

Our methodological results are encouraging for the internal validity of the interval regression approach, and future work will address the question of external validity. In particular, we aim to investigate whether the Canadian HUI values or threshold values obtained from a European survey can be used to scale interval regressions applied to other surveys, such as the European Community Household Panel (ECHP), that only include self-assessed health and do not measure HUI directly.

References

Cutler, D., Richardson, E., 1997. Measuring the health of the United States Population. *Brookings Papers on Economic Activity, Microeconomics*, 217-271.

Feeny, D., Furlong, W., Boyle, M., Torrance, G., 1995. Multi-attribute health status classification systems: Health Utilities Index. *Pharmacoeconomics* 7 (6), 490-502.

Furlong, W., Feeny, D., Torrance, G.W., Goldsmith, C., DePauw, S., Zhu, Z., Denton, M., and Boyle, M., 1998. *Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report*, McMaster University, Centre for Health Economics and Policy Analysis Working Paper No. 98-11.

Gakidou, E, Murray, C., Frenk, J., 2000. Defining and measuring health inequality. *Bulletin of the World Health Organisation* 78(1), 42-52.

Gerdtham, U-G., Johanneson, M., 2000. Income-related inequality in life-years and quality-adjusted life-years. *Journal of Health Economics* 19(6), 1007-1026.

Gerdtham, U-G., Johanneson, M., Lundberg, L., Isacson, D., 1999, A note on validating Wagstaff and van Doorslaer's health measure in the analysis of inequalities in health, *Journal of Health Economics*, 18, 117-124

Groot, W., 2000. Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics* 19(3), 403-420.

Humphries, K and E van Doorslaer, 2000, Income-related inequalities in health in Canada, *Social Science and Medicine*, 50, 663-671

Idler, E. L. and Y. Benyamini (1997). "Self-rated health and mortality: a review of twenty-seven community studies." *Journal of Health and Social Behavior* 38(1): 21-37.

Jones, A.M., 2000. Health Econometrics, in A.J.Culyer and J.P.Newhouse (eds.) *Handbook of Health Economics.* Elsevier, Amsterdam.

Kakwani, N.C., 1980. *Income Inequality and Poverty. Methods of Estimation and Policy Implications.* Oxford University Press, Oxford.

Kakwani, N.C., Wagstaff, A., van Doorslaer, E., 1994. Socioeconomic inequalities in health: measurement, computation and statistical inference. *Journal of Econometrics*, 77(1), 87-104.

Kerkhofs, M., Lindeboom, M., 1995, Subjective health measures and state dependent reporting errors. *Health Economics* 4, 221-235.

Le Grand, J., 1989. An international comparison of ages-at-death. in J. Fox (ed.) *Health inequalities in European countries.* Gower, Aldershot.

Murray, CJL, A Tandon, J Salomon and CD Mathers, 2001, *Enhancing cross-population comparability of survey results*, GPE Discussion Paper Nr 35, WHO/EIP, Geneva.

Lerman, R.I., Yitzhaki, S., 1989. Improving the accuracy of estimates of Gini coefficients. *Journal of Econometrics* 42, 43-47.

Sadana, R, CD Mathers, AD Lopez, CJL Murray and K Iburg, 2000, *Comparative analysis of more than 50 household surveys on health status*, GPE Discussion Paper No 15, EIP/GPE/EBD, World Health organisation, Geneva.

Tambay J-L., Catlin, G., 1995. Sample Design of the National Population Health Survey. *Health Reports* 7(1), 29-38.

Torrance, G.W., Furlong, W., Feeny, D., Boyle, M., 1995. Multi-attribute preference functions: Health Utilities Index. *Pharmacoeconomics* 7(6), 503-520.

Torrance, G.W., Feeny, D., Furlong W.J., Barr, R.D., Zhang, Y., Wang, Q., 1996. Multiattribute Utility Function for a Comprehensive Health Status Classification System. *Medical Care* 34(7), 702-722.

Van Doorslaer, E, Koolman, X., 2000, *Income-related inequalities in health in Europe: evidence from the European Community Household Panel*, Ecuity II Project, Working Paper #1, Erasmus University, Rotterdam.

Van Doorslaer, E., Wagstaff, A., Bleichrodt, H., *et al*, 1997. Income-related inequalities in health: some international comparisons. *Journal of Health Economics* 16, 93-112.

Wagstaff, A., van Doorslaer, E., 1994. Measuring inequalities in health in the presence of multiple-category morbidity indicators. *Health Economics 3*, 281-291.

Wagstaff, A., van Doorslaer, E., Paci, P., 1989. Equity in the finance and delivery of health care: some tentative cross-country comparisons. *Oxford Review of Economic Policy* 5(1), 89-112.

Wagstaff, A., Paci, P., van Doorslaer, E., 1991. On the measurement of inequalities in health. *Social Science and Medicine 33*, 545-557.

Wagstaff, A, van Doorslaer, E., Watanabe, N., 2002. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam, *Journal of Econometrics*, (forthcoming).

Fig. 1. HUI interval boundaries of SAH levels derived from the empirical cumulative distribution for the Canadian NPHS 1994 sample.

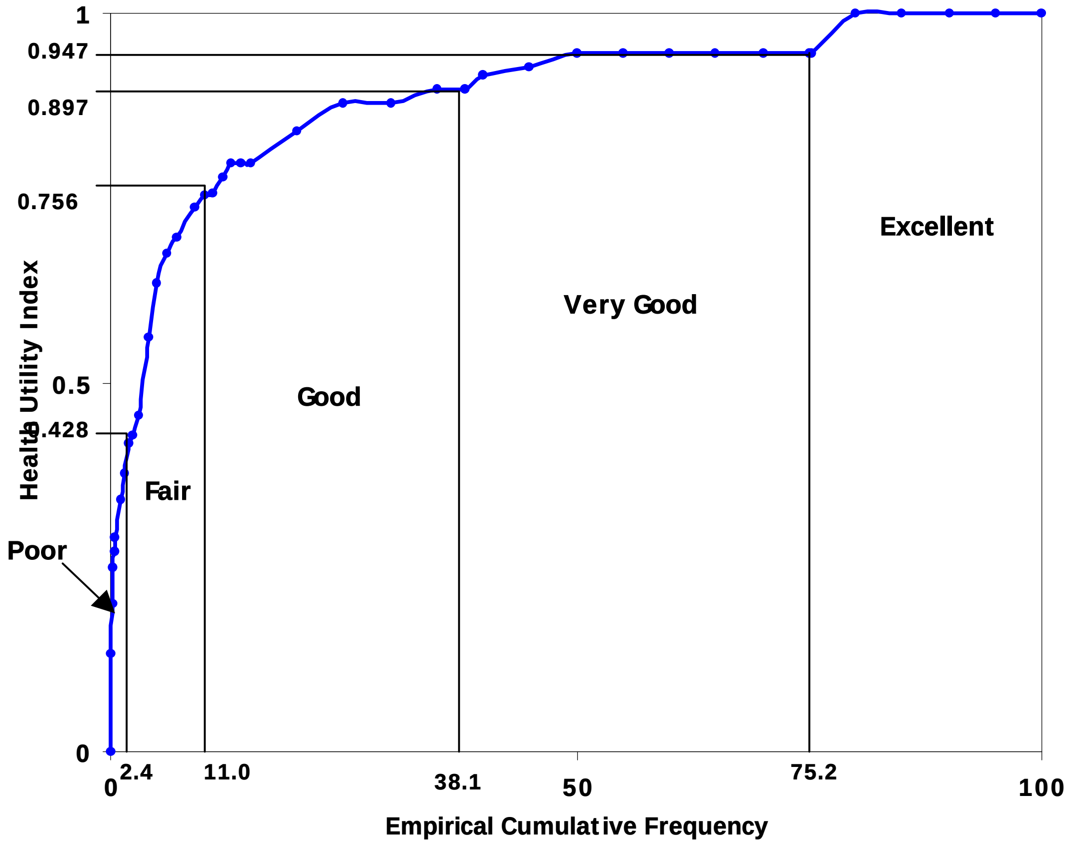

Fig 2: Health concentration curves

(as % deviation from diagonal)

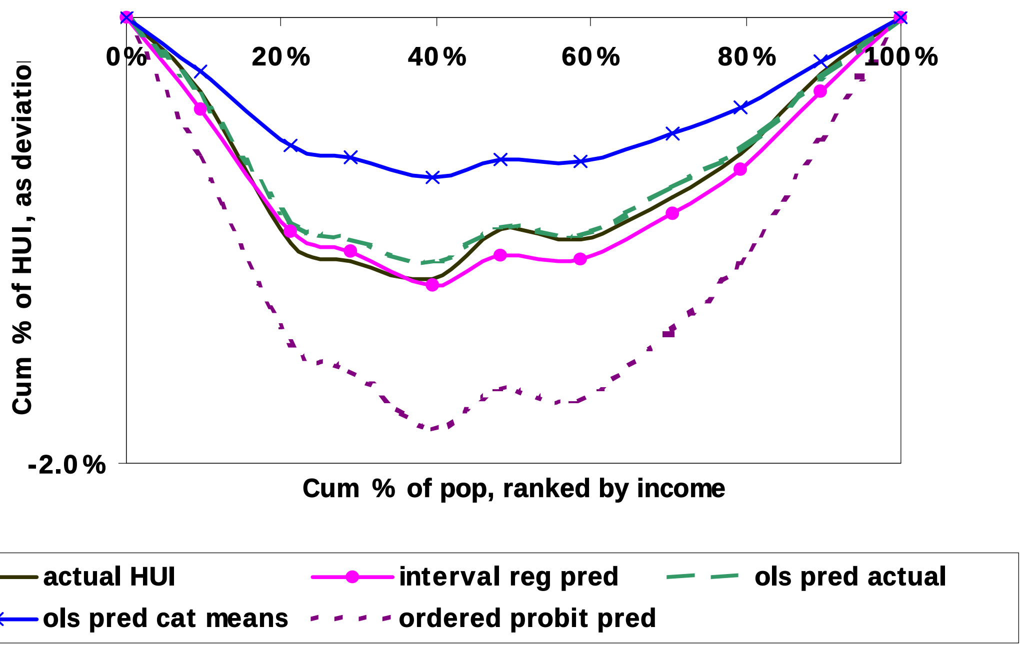

Table 1: Characteristics of the distribution of HUI

Self-assessed health	Number of respondents	Sample proportion	Mean HUI	Std. Dev	Upper bound of interval
Poor	374	2.4%	0.557	0.241	0.428
Fair	1343	8.6%	0.758	0.184	0.756
Good	4198	27.0%	0.876	0.127	0.897
Very Good	5771	37.2%	0.923	0.082	0.947
Excellent	3853	24.8%	0.945	0.064	1.000
Total	15539	100.0%	0.893	0.132	

Table 2: Descriptive statistics for HUI by category of SAH, total and subgroups

	Total	by demographic group				by disability status	
		M18-44	F18-44	M45+	F45+	Disabled	Non-disab
N	15539	4438	4478	3154	3469	433	15106
Mean HUI							
Poor	0.557	0.577	0.581	0.582	0.522	0.481	0.603
Fair	0.758	0.797	0.763	0.767	0.736	0.636	0.768
Good	0.876	0.894	0.896	0.865	0.849	0.700	0.876
Very good	0.923	0.938	0.929	0.917	0.893	0.830	0.922
Excellent	0.945	0.955	0.952	0.927	0.923	0.814	0.944
Cum % of sample							
Poor	2.4%	0.8%	1.1%	4.4%	4.4%	32.2%	1.6%
Fair	11.0%	4.4%	6.5%	17.6%	19.5%	67.7%	9.4%
Good	38.1%	27.8%	31.0%	48.3%	51.0%	87.6%	36.6%
Very good	75.2%	69.7%	71.0%	81.4%	82.0%	97.8%	74.6%
Excellent	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Upper bound of interval							
Poor	0.428	0.394	0.418	0.454	0.428	0.435	0.428
Fair	0.756	0.737	0.745	0.796	0.753	0.753	0.756
Good	0.897	0.917	0.897	0.897	0.897	0.947	0.897
Very good	0.947	1	1	0.947	0.947	1	0.947
Excellent	1	1	1	1	1	1	1

Table 3a: Descriptive statistics actual and predicted HUI values (n=15539)

Variable	Mean	Std. Dev.	Min	P(25)	P(50)	P(75)	Max	Conc Index	Gini
Actual HUI	0.8926	0.1324	0.0310	0.8770	0.9470	0.9470	1.0000	0.0141	0.0675
HUI categ means	0.8930	0.0730	0.5585	0.8764	0.9230	0.9230	0.9452	0.0097	0.0175
OLS actual	0.8926	0.0648	0.4578	0.8764	0.9129	0.9313	0.9666	0.0135	0.0326
OLS categ means	0.8930	0.0354	0.6870	0.8822	0.9041	0.9148	0.9377	0.0090	0.0185
Interval regression	0.8876	0.0595	0.5379	0.8704	0.9056	0.9234	0.9627	0.0151	0.0309
Ordered probit	1.1143	0.4967	-1.4544	0.8780	1.2405	1.4534	1.9641	0.1196	0.2337
Rescaled ord prob 1	0.7514	0.1453	0.0000	0.6823	0.7883	0.8506	1.0000	0.0519	0.1014
Rescaled ord prob 2	0.8401	0.0739	0.4578	0.8050	0.8589	0.8906	0.9666	0.0236	0.0461

Table 3b: Predictions, conditional on category of SAH

SAH = poor (n=374)	**Mean**	**Std. Dev.**	**Min**	**P(25)**	**P(50)**	**P(75)**	**Max**
Actual	0.5569	0.2411	0.0310	0.3620	0.5110	0.7560	1.0000
OLS actual	0.3907	0.0118	0.3545	0.3797	0.3972	0.4003	0.4043
OLS categ means	0.4177	0.0017	0.4138	0.4160	0.4185	0.4191	0.4201
Interval regression	0.4010	0.0083	0.3808	0.3929	0.4053	0.4075	0.4109
Ordered probit	-1.6928	0.1339	-2.0170	-1.8052	-1.6507	-1.5901	-1.4901
Rescaled ord probit 1	-0.0698	0.0392	-0.1646	-0.1026	-0.0574	-0.0397	-0.0105
Rescaled ord probit 2	0.4223	0.0199	0.3741	0.4056	0.4286	0.4376	0.4525
SAH = fair (n=1343)							
Actual	0.7577	0.1840	0.0770	0.6930	0.7960	0.8970	1.0000
OLS actual	0.6938	0.0353	0.5305	0.6927	0.7066	0.7153	0.7237
OLS categ means	0.7411	0.0133	0.6757	0.7409	0.7452	0.7488	0.7524
Interval regression	0.7070	0.0354	0.5591	0.7079	0.7184	0.7269	0.7354
Ordered probit	-0.6378	0.0418	-0.7842	-0.6574	-0.6329	-0.6061	-0.5627
Rescaled ord probit 1	0.2389	0.0122	0.1960	0.2331	0.2403	0.2481	0.2608
Rescaled ord probit 2	0.5794	0.0062	0.5576	0.5764	0.5801	0.5841	0.5905
SAH = good (n=4198)							
Actual	0.8764	0.1266	0.1110	0.8300	0.9170	0.9470	1.0000
OLS actual	0.8398	0.0053	0.8057	0.8379	0.8413	0.8433	0.8469
OLS categ means	0.8492	0.0082	0.8026	0.8457	0.8515	0.8546	0.8601
Interval regression	0.8410	0.0066	0.8040	0.8380	0.8427	0.8453	0.8507
Ordered probit	0.3364	0.0406	0.1391	0.3130	0.3442	0.3659	0.4143
Rescaled ord probit 1	0.5238	0.0119	0.4661	0.5170	0.5261	0.5325	0.5466
Rescaled ord probit 2	0.7243	0.0060	0.6950	0.7209	0.7255	0.7287	0.7359
SAH = very good (n=5771)							
Actual	0.9229	0.0818	0.2690	0.8970	0.9470	0.9680	1.0000
OLS actual	0.9317	0.0014	0.9201	0.9313	0.9321	0.9325	0.9335
OLS categ means	0.9294	0.0023	0.9140	0.9285	0.9300	0.9309	0.9330
Interval regression	0.9311	0.0018	0.9179	0.9304	0.9315	0.9322	0.9338
Ordered probit	1.3233	0.0354	1.1228	1.3048	1.3310	1.3482	1.3914
Rescaled ord probit 1	0.8125	0.0103	0.7539	0.8071	0.8148	0.8198	0.8325
Rescaled ord probit 2	0.8712	0.0053	0.8414	0.8685	0.8724	0.8749	0.8814
SAH = Excellent (n=3853)							
Actual	0.9447	0.0644	0.3710	0.9260	0.9470	1.0000	1.0000
OLS actual	0.9836	0.0002	0.9816	0.9835	0.9836	0.9837	0.9839
OLS categ means	0.9824	0.0004	0.9791	0.9822	0.9825	0.9826	0.9830
Interval regression	0.9833	0.0003	0.9807	0.9832	0.9834	0.9835	0.9838
Ordered probit	2.5021	0.0818	2.1782	2.4553	2.5109	2.5596	2.6875
Rescaled ord probit 1	1.1574	0.0239	1.0626	1.1437	1.1600	1.1742	1.2116
Rescaled ord probit 2	1.0467	0.0122	0.9985	1.0397	1.0480	1.0552	1.0743

Table 4: Regression and decomposition results

	Observed HUI		OLS predicted using actual HUI			OLS predicted using HUI means by SAH			Interval regression predicted			(rescaled) ordered probit		
Conc Ind C	0.01408				0.0135			0.0090			0.0151			0.0236
St error C	0.00092				0.00043			0.00023			0.00038			0.00048
	Mean	CI	Coef.	t	% contrib	Coef.	t	% contrib	Coef.	z	% contrib	Coef.	z	% contrib
lincome	9.9828	0.0377	0.0095	4.92	29.79%	0.0090	8.25	41.95%	0.0153	8.72	42.82%	0.0245	8.98	46.56%
educ1	0.0767	-0.3698	-0.0541	-6.43	12.73%	-0.0336	-8.18	11.88%	-0.0562	-8.37	11.88%	-0.0834	-9.64	11.92%
educ2	0.1835	-0.2248	-0.0170	-4.55	5.83%	-0.0163	-7.51	8.39%	-0.0285	-8.25	8.77%	-0.0589	-9.91	12.25%
educ3	0.1606	-0.0138	-0.0059	-1.68	0.11%	-0.0079	-3.81	0.22%	-0.0148	-4.49	0.24%	-0.0352	-5.93	0.39%
educ4	0.3535	0.0305	-0.0103	-3.76	-0.92%	-0.0069	-4.32	-0.93%	-0.0127	-4.97	-1.02%	-0.0314	-6.30	-1.71%
househ	0.1451	-0.2205	-0.0208	-4.49	5.52%	-0.0093	-4.24	3.70%	-0.0146	-4.17	3.48%	-0.0265	-4.56	4.27%
student	0.0657	-0.1254	-0.0021	-0.43	0.14%	-0.0052	-1.74	0.54%	-0.0070	-1.44	0.43%	-0.0063	-0.66	0.26%
disabled	0.0279	-0.3187	-0.2895	-19.66	21.36%	-0.1540	-18.14	17.05%	-0.2664	-15.94	17.63%	-0.2543	-20.73	11.39%
unemploy	0.0342	-0.3292	-0.0123	-2.22	1.15%	-0.0043	-1.35	0.60%	-0.0075	-1.43	0.63%	-0.0168	-1.72	0.96%
retired	0.1585	-0.2063	-0.0331	-4.79	8.99%	-0.0226	-5.74	9.21%	-0.0359	-5.75	8.74%	-0.0532	-6.28	8.77%
other	0.0128	-0.2390	-0.0272	-2.13	0.69%	-0.0217	-2.32	0.82%	-0.0355	-2.47	0.81%	-0.0408	-2.62	0.63%
married	0.6593	0.0646	0.0087	2.43	3.06%	0.0011	0.55	0.59%	0.0022	0.69	0.71%	0.0058	1.10	1.25%
div_wid	0.1339	-0.2381	-0.0069	-1.30	1.82%	-0.0026	-0.87	1.03%	-0.0016	-0.34	0.39%	0.0034	0.49	-0.54%
m20_24	0.0389	0.0077	0.0093	1.04	0.02%	0.0043	0.58	0.02%	0.0071	0.65	0.02%	0.0211	1.00	0.03%
m25_29	0.0503	0.0515	-0.0026	-0.28	-0.06%	0.0001	0.01	0.00%	0.0000	0.00	0.00%	0.0066	0.32	0.09%
m30_34	0.0588	0.0852	-0.0099	-0.99	-0.41%	-0.0055	-0.75	-0.34%	-0.0092	-0.83	-0.34%	-0.0099	-0.47	-0.25%
m35_39	0.0622	0.0431	-0.0044	-0.45	-0.10%	-0.0055	-0.74	-0.18%	-0.0115	-1.02	-0.23%	-0.0185	-0.89	-0.25%
m40_44	0.0534	0.1048	-0.0134	-1.41	-0.62%	-0.0072	-0.97	-0.50%	-0.0129	-1.16	-0.54%	-0.0212	-1.01	-0.60%
m45_49	0.0480	0.2388	-0.0235	-2.42	-2.24%	-0.0118	-1.53	-1.68%	-0.0205	-1.77	-1.75%	-0.0308	-1.43	-1.78%
m50_54	0.0367	0.1949	-0.0307	-2.66	-1.82%	-0.0192	-2.31	-1.71%	-0.0314	-2.53	-1.68%	-0.0506	-2.30	-1.83%
m55_59	0.0281	0.1411	-0.0363	-3.33	-1.19%	-0.0231	-2.84	-1.14%	-0.0384	-3.13	-1.13%	-0.0624	-2.83	-1.24%
m60_64	0.0253	-0.0398	-0.0359	-2.62	0.30%	-0.0227	-2.35	0.28%	-0.0397	-2.70	0.30%	-0.0632	-2.75	0.32%
m65_69	0.0240	-0.1274	-0.0304	-2.34	0.77%	-0.0302	-2.92	1.15%	-0.0500	-3.14	1.14%	-0.0693	-2.89	1.07%
m70_74	0.0185	-0.1141	-0.0521	-3.19	0.91%	-0.0278	-2.76	0.73%	-0.0441	-2.87	0.69%	-0.0668	-2.76	0.71%
m75_79	0.0121	-0.3042	-0.0439	-2.81	1.34%	-0.0311	-2.81	1.43%	-0.0515	-3.06	1.41%	-0.0740	-2.97	1.37%
m80_	0.0104	-0.2490	-0.1177	-5.48	2.53%	-0.0428	-3.22	1.38%	-0.0681	-3.23	1.32%	-0.0830	-2.90	1.09%
f15_19	0.0162	-0.0727	-0.0017	-0.17	0.02%	0.0029	0.37	-0.04%	0.0033	0.28	-0.03%	0.0091	0.40	-0.05%
f20_24	0.0429	-0.1684	-0.0090	-0.94	0.54%	-0.0005	-0.07	0.05%	-0.0016	-0.14	0.09%	-0.0015	-0.07	0.05%
f25_29	0.0501	-0.0367	-0.0068	-0.71	0.10%	-0.0073	-0.93	0.17%	-0.0111	-0.95	0.15%	-0.0077	-0.36	0.07%
f30_34	0.0656	-0.0406	-0.0070	-0.74	0.15%	-0.0040	-0.55	0.13%	-0.0071	-0.64	0.14%	-0.0041	-0.20	0.06%
f35_39	0.0602	-0.0151	-0.0120	-1.26	0.09%	-0.0082	-1.10	0.09%	-0.0136	-1.21	0.09%	-0.0177	-0.84	0.08%
f40_44	0.0532	0.0825	-0.0220	-2.12	-0.80%	-0.0123	-1.62	-0.67%	-0.0215	-1.88	-0.70%	-0.0331	-1.57	-0.73%

f45_49	0.0435	0.1578	-0.0427	-4.18	-2.43%	-0.0181	-2.23	-1.55%	-0.0291	-2.38	-1.49%	-0.0421	-1.95	-1.46%
f50_54	0.0349	0.1220	-0.0474	-3.77	-1.67%	-0.0171	-2.15	-0.91%	-0.0305	-2.53	-0.97%	-0.0525	-2.42	-1.12%
f55_59	0.0313	0.0515	-0.0460	-3.99	-0.62%	-0.0261	-3.08	-0.52%	-0.0421	-3.25	-0.51%	-0.0584	-2.60	-0.47%
f60_64	0.0287	-0.1697	-0.0319	-2.69	1.29%	-0.0160	-1.80	0.97%	-0.0267	-1.94	0.97%	-0.0462	-2.03	1.14%
f65_69	0.0265	-0.2164	-0.0500	-3.62	2.38%	-0.0208	-2.24	1.48%	-0.0362	-2.61	1.55%	-0.0591	-2.54	1.71%
f70_74	0.0253	-0.2334	-0.0525	-3.77	2.58%	-0.0201	-2.21	1.48%	-0.0333	-2.38	1.47%	-0.0542	-2.28	1.62%
f75_79	0.0169	-0.2690	-0.0707	-4.53	2.67%	-0.0367	-3.51	2.08%	-0.0572	-3.52	1.94%	-0.0741	-3.05	1.70%
f80_	0.0161	-0.3669	-0.1223	-6.90	5.99%	-0.0372	-3.40	2.74%	-0.0588	-3.40	2.59%	-0.0768	-3.01	2.28%
_const			0.8455	39.43		0.8346	67.29		0.7870	39.38				

Table 5: Health Gini index: decomposition results

	Observed HUI		OLS predicted using actual HUI			OLS predicted using HUI means by SAH			Interval regression predicted			(rescaled) Ordered Probit predicted		
Gini Index *G*	0.0675			0.0326			0.0185			0.0309			0.0461	
St error *G*	0.00080			0.00041			0.00021			0.00037			0.00032	
C/G (%)	20.9%			41.4%			48.5%			49.0%			51.2%	
	Mean	CI hui	CI prhui1	Contrib	% contrib	CI prhui2	Contrib	% contrib	CI predhui	Contrib	% contrib	CI prhui3	Contrib	% contrib
lincome	9.9828	0.0049	0.0161	0.00172	5.3%	0.0199	0.00199	10.7%	0.0202	0.00347	11.2%	0.0199	0.00580	12.6%
educ1	0.0767	-0.3235	-0.7951	0.00370	11.3%	-0.7959	0.00230	12.4%	-0.7993	0.00388	12.6%	-0.7753	0.00590	12.8%
educ2	0.1835	-0.0917	-0.3197	0.00112	3.4%	-0.4117	0.00138	7.5%	-0.4181	0.00247	8.0%	-0.4454	0.00573	12.4%
educ3	0.1606	0.0458	0.0962	-0.00010	-0.3%	0.0149	-0.00002	-0.1%	0.0012	0.00000	0.0%	-0.0259	0.00017	0.4%
educ4	0.3535	0.0264	0.0618	-0.00025	-0.8%	0.0984	-0.00027	-1.4%	0.0952	-0.00048	-1.6%	0.0731	-0.00097	-2.1%
househ	0.1451	-0.0559	-0.3541	0.00120	3.7%	-0.3049	0.00046	2.5%	-0.2960	0.00071	2.3%	-0.2987	0.00137	3.0%
student	0.0657	0.1292	0.3642	-0.00006	-0.2%	0.2824	-0.00011	-0.6%	0.3045	-0.00016	-0.5%	0.3297	-0.00016	-0.4%
disabled	0.0279	-0.6809	-0.9722	0.00879	27.0%	-0.9722	0.00468	25.2%	-0.9722	0.00814	26.4%	-0.9721	0.00821	17.8%
unemploy	0.0342	-0.0147	-0.0895	0.00004	0.1%	-0.0103	0.00000	0.0%	-0.0307	0.00001	0.0%	-0.0745	0.00005	0.1%
retired	0.1585	-0.2791	-0.7272	0.00428	13.1%	-0.7420	0.00298	16.1%	-0.7380	0.00473	15.3%	-0.7289	0.00732	15.9%
other	0.0128	-0.0783	-0.4504	0.00018	0.5%	-0.5316	0.00016	0.9%	-0.5336	0.00027	0.9%	-0.4196	0.00026	0.6%
married	0.6593	0.0225	0.0449	0.00029	0.9%	0.0116	0.00001	0.1%	0.0088	0.00001	0.0%	0.0082	0.00004	0.1%
div_wid	0.1339	-0.2003	-0.5339	0.00055	1.7%	-0.4446	0.00017	0.9%	-0.4216	0.00010	0.3%	-0.3922	-0.00021	-0.5%
m20_24	0.0389	0.2195	0.7005	0.00028	0.9%	0.6033	0.00011	0.6%	0.6043	0.00019	0.6%	0.6187	0.00060	1.3%
m25_29	0.0503	0.1513	0.5510	-0.00008	-0.2%	0.5700	0.00000	0.0%	0.5642	0.00000	0.0%	0.5569	0.00022	0.5%
m30_34	0.0588	0.1730	0.4129	-0.00027	-0.8%	0.3886	-0.00014	-0.8%	0.3909	-0.00024	-0.8%	0.3914	-0.00027	-0.6%
m35_39	0.0622	0.2020	0.5361	-0.00016	-0.5%	0.3632	-0.00014	-0.7%	0.3138	-0.00025	-0.8%	0.2643	-0.00036	-0.8%
m40_44	0.0534	0.0906	0.3087	-0.00025	-0.8%	0.3023	-0.00013	-0.7%	0.2924	-0.00023	-0.7%	0.2477	-0.00033	-0.7%
m45_49	0.0480	-0.0122	0.0680	-0.00009	-0.3%	0.1464	-0.00009	-0.5%	0.1367	-0.00015	-0.5%	0.1362	-0.00024	-0.5%
m50_54	0.0367	-0.0810	-0.1634	0.00021	0.6%	-0.2037	0.00016	0.9%	-0.1882	0.00024	0.8%	-0.1887	0.00042	0.9%
m55_59	0.0281	-0.1347	-0.3214	0.00037	1.1%	-0.3676	0.00027	1.4%	-0.3632	0.00044	1.4%	-0.3654	0.00076	1.7%

m60_64	0.0253	-0.1344	-0.4991	0.00051	1.6%	-0.5526	0.00035	1.9%	-0.5687	0.00064	2.1%	-0.5704	0.00108	2.4%
m65_69	0.0240	-0.1776	-0.5128	0.00042	1.3%	-0.7096	0.00057	3.1%	-0.7102	0.00096	3.1%	-0.6770	0.00134	2.9%
m70_74	0.0185	-0.2554	-0.7282	0.00078	2.4%	-0.7308	0.00042	2.3%	-0.7164	0.00066	2.1%	-0.7056	0.00104	2.2%
m75_79	0.0121	-0.2842	-0.6990	0.00042	1.3%	-0.7866	0.00033	1.8%	-0.7881	0.00055	1.8%	-0.7827	0.00083	1.8%
m80_	0.0104	-0.4542	-0.9076	0.00125	3.8%	-0.8636	0.00043	2.3%	-0.8583	0.00069	2.2%	-0.8166	0.00084	1.8%
f15_19	0.0162	0.0835	0.2933	-0.00001	0.0%	0.3699	0.00002	0.1%	0.3449	0.00002	0.1%	0.3207	0.00006	0.1%
f20_24	0.0429	0.1028	0.1691	-0.00007	-0.2%	0.3344	-0.00001	0.0%	0.3258	-0.00002	-0.1%	0.2726	-0.00002	0.0%
f25_29	0.0501	0.1345	0.3401	-0.00013	-0.4%	0.2073	-0.00009	-0.5%	0.2401	-0.00015	-0.5%	0.3133	-0.00014	-0.3%
f30_34	0.0656	0.1450	0.3211	-0.00017	-0.5%	0.3063	-0.00009	-0.5%	0.3083	-0.00016	-0.5%	0.3361	-0.00011	-0.2%
f35_39	0.0602	0.1064	0.1978	-0.00016	-0.5%	0.1405	-0.00008	-0.4%	0.1530	-0.00014	-0.5%	0.1606	-0.00020	-0.4%
f40_44	0.0532	0.0726	-0.0054	0.00001	0.0%	0.0346	-0.00003	-0.1%	0.0295	-0.00004	-0.1%	0.0238	-0.00005	-0.1%
f45_49	0.0435	-0.1449	-0.3712	0.00077	2.4%	-0.1693	0.00015	0.8%	-0.1373	0.00020	0.6%	-0.0900	0.00020	0.4%
f50_54	0.0349	-0.1178	-0.4754	0.00088	2.7%	-0.2212	0.00015	0.8%	-0.2400	0.00029	0.9%	-0.2718	0.00059	1.3%
f55_59	0.0313	-0.1680	-0.5124	0.00083	2.5%	-0.4951	0.00045	2.4%	-0.4763	0.00071	2.3%	-0.4153	0.00090	2.0%
f60_64	0.0287	-0.2079	-0.5143	0.00053	1.6%	-0.4825	0.00025	1.3%	-0.4754	0.00041	1.3%	-0.5051	0.00080	1.7%
f65_69	0.0265	-0.2921	-0.7065	0.00105	3.2%	-0.6264	0.00039	2.1%	-0.6342	0.00069	2.2%	-0.6515	0.00121	2.6%
f70_74	0.0253	-0.2845	-0.7488	0.00112	3.4%	-0.6652	0.00038	2.0%	-0.6598	0.00063	2.0%	-0.6720	0.00110	2.4%
f75_79	0.0169	-0.3357	-0.8311	0.00111	3.4%	-0.8237	0.00057	3.1%	-0.8086	0.00088	2.9%	-0.7722	0.00115	2.5%
f80_	0.0161	-0.4968	-0.9206	0.00203	6.2%	-0.8541	0.00057	3.1%	-0.8444	0.00090	2.9%	-0.8182	0.00120	2.6%