A LONGITUDINAL ANALYSIS OF INCOME-RELATED HEALTH INEQUALITY IN AUSTRALIA♠

Jason D. Brandrup and Michael A. Kortt ♣
Research and Analysis Branch
Department of Families, Community Services and Indigenous Affairs

Correspondence:
Michael A. Kortt, PhD
Section Manager
HILDA Survey and Research Projects
Research and Analysis Branch
Department of Families, Community Services and Indigenous Affairs
Box 7788 Canberra Mail Centre
ACT 2610

Tel: (02) 6244 1814
Fax: (02) 6244 6589
E-mail: <michael.kortt@facsia.gov.au>

This version: 2 July 2007

♠ Prepared for the HILDA Survey Research Conference, University of Melbourne, 19–20 July 2007. Thanks to Andrew Leigh, Bob Breunig, Philip Clarke and our colleagues, Andrew Whitecross, Rob Bray, Ibolya Losoncz, Peng Yu and Annette Neuendorf, FaCSIA, for comments/discussion on earlier versions of this paper.

♣ This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey (Release 5.1). The HILDA Project was initiated and is funded by the Australian Government Department of Families, Community Services and Indigenous Affairs (FaCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (MIAESR). The findings and views reported in this paper; however, are the views of the authors and should not be attributed to either FaCSIA or MIAESR. Moreover, the opinions, comments and/or analysis expressed in this document are those of the authors and do not necessarily represent the views of the Minister for Families, Community Services and Indigenous Affairs or the Australian Government Department of Families, Community Services and Indigenous Affairs, and cannot be taken in any way as expressions of Government policy.

Abstract

The traditional approach to measuring income-related health inequalities has predominantly relied on cross-sectional data to estimate measures of inequality such as concentration indices. Jones and López-Nicolás[1], however, have demonstrated that when there are differences in the health between those individuals who are upwardly or downwardly income mobile, longitudinal income-related inequality measures can show a greater degree of inequality compared to measures calculated using cross-sectional data.

Subsequently, Jones and López-Nicolás[1] developed a *health-related mobility index* to measure the difference between cross-sectional and longitudinal income-related health inequality concentration indices, which can be decomposed into the contribution of different regressors.

We replicated the method used by Jones and López-Nicolás. [1] Specifically, we developed a *health-related mobility index* for Australia using utilities from the SF-6D, a summary preference-based measure of health derived from the SF-36 questionnaire from the first five waves (2001–2005) of the Household, Income and Labour Dynamics in Australia (HILDA) Survey. We also decomposed our mobility index to examine the contribution of different regressors.

Keywords: concentration indices; decomposition; HILDA; income-related health inequalities; mobility; SF-6D; SF-36.

1. Introduction

Traditionally, the measurement of health inequalities has predominantly relied on cross-sectional data. Health economists, for example, have typically used concentration indices to measure socioeconomic health inequalities, with socioeconomic status usually defined in terms of income. As income is often used as a broad measure of socioeconomic status, concentration indices are more specifically used to quantify income-related health inequalities.[2-5] Using this approach, it is possible to incorporate a regression model for health to decompose the concentration index into the contribution of different explanatory variables.[6]

With the rise of international longitudinal data collections, researchers have started to examine the dynamics of health and its relationship to a range of socioeconomic characteristics.[1, 7] Jones and López-Nicolás [1], for instance, have demonstrated that when there are differences in health between those individuals who are upwardly or downwardly income mobile, longitudinal income-related inequality measures can show a greater degree of inequality compared to measures calculated using cross sectional data. The authors developed a *health-related mobility index* (which can also be decomposed) using the first nine waves of the British Household Panel Survey to measure the differences between cross-sectional and longitudinal income-related health inequality concentration indices.

An advantage of using the Jones and López-Nicolás[1] approach is that it can potentially uncover and identify important characteristics of income-related health inequalities that cannot be revealed by cross-sectional data (e.g., the positive association between health and income over time). By using longitudinal data, it is

possible to gain a more detailed insight into how health inequalities are changing over time and what factors might be associated with such a change.

The aim of this paper is to contribute to the literature by using the Jones and López-Nicolás approach to develop a *health-related mobility index* for Australia using utilities estimated via the SF-6D, a summary preference-based measure of health derived from the SF-36 questionnaire for the first five waves (2001-2005) of the Household, Income and Labour Dynamics in Australia (HILDA) Survey. We illustrate how the mobility index can be decomposed by using a simple linear regression model and a fixed-effects model.

The paper itself is organised as follows. In section 2 we set out the methods used to measure and decompose both cross-sectional and longitudinal concentration indices. Section 3 describes the data and variables used and section 4 provides the results of our analysis. We end with a discussion and conclusions.

2. Methods

Measurement of health

Health information is often available at an ordinal level.[8] For example, one of the most common indicators of self-assessed health (SAH), which is frequently included in population health surveys, is the question: 'In general, would you say your health is: excellent; very good; good; fair; poor?'

While this categorical variable has been shown to be a good predictor of health outcomes such as mortality[9] it does, however, pose a challenge for health inequality

measurement, as the calculation of the concentration index requires that the health variable is either in a dichotomous or continuous format.[3] Dichotomising a variable into, say, 'good health' and 'poor health' is less than ideal because it results in a loss of information and the choice of cut-off point is arbitrary. An alternative approach is to transform the categorical variable into a continuous variable. A variety of transformation options are available.[8]

For example, Wagstaff and Van Doorslaer[3] assumed that the underlying categorical distribution of the SAH question was a latent, continuous but unobservable health variable with a lognormal distribution. This allows for the 'scoring' of SAH categories using the midpoints of the intervals corresponding to the lognormal distribution. Another approach is to estimate an ordered logistic or probit regression models using SAH categories as the dependent variable and then re-scale the underlying latent variable to compute 'weights' for health between 0 and 1.[8] If information on the scaling is available (e.g., when the lower and upper limits of the intervals are known), then interval regression provides an alternative approach to the ordered logistic and probit regression models. [8]

To address the health measurement issue we made use of the SF-36, which is a generic quality-of-life instrument that was administered across all five waves of HILDA. The SF-36 health survey is a questionnaire that is used to evaluate patient health status across eight separate dimensions: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health.[10] Each dimension is scored from 0 to 100 where higher scores correspond

to a better health status. These scores, however, do not provide an overall preference-based measure of health outcome.[11, 12]

The preference-based measurement of health outcomes has, however, been advanced by the development of several algorithms to convert item responses to generic health surveys into health state utility values.[13-17] In this study, we employ utilities estimated using the SF-6D[11], a summary preference-based measure of health derived from the SF-36, to estimate the health status of people participating in HILDA. The advantage of this approach is that it allows us to generate a continuous measure of health status.

Measurement of inequality

Suppose we have a continuous cardinal measure of health, y_i. The concentration curve, L, plots the cumulative proportion of individuals ranked by income (X-axis) against the cumulative proportion of health (Y-axis). If the concentration curve, L, coincides with the diagonal, then everyone, regardless of income, reports the same level of health. If, however, the curve lies below (above) the diagonal then health inequalities exist and richer (poorer) people have a better health status (Figure 1). The concentration index, C, is defined as twice the area between the concentration curve, L, and the diagonal. It measures the degree of inequality or, in other words, how far the concentration curve, L, diverges from the diagonal. When the concentration index coincides with the diagonal, the concentration index is zero. The concentration index will be positive (negative) if the curve lies above (below) the diagonal, indicating that inequality favours the poor (rich). The minimum and maximum values of the concentration index are -1 and 1+, respectively.

The cross-sectional concentration index, *CX*, can be computed using a modified version of the convenience covariance formula that incorporates cross-sectional sample weights[5]:

$$CX = \frac{2}{\mu} w\mathrm{cov}_x(y_i, R_i) \ldots\ldots\ldots\ldots (1)$$

Where:

y_i is the measure of health status, $w\mathrm{cov}_x$ is the cross-sectional weighted covariance, $\mu = \sum_{i=1}^{N} w_i y_i$ is the weighted mean health status of the sample, N is the sample size, w_i is the sample weight of individual i (with the summation of w_i equal to N), and R_i is the weighted fractional income rank.

Cross-sectional concentration indices, however, may miss the positive association between income and health that individuals may experience over time.[1] To capture this association in the measurement of inequality over time, a longitudinal concentration index, *CL*, can also be computed in terms of the covariance between health status and relative fractional rank for the same individual incorporating longitudinal sample weights:

$$CL = \frac{2}{Y^T} w\mathrm{cov}_l(y_i^T, R_i^T) \ldots\ldots\ldots\ldots (2)$$

Where Y^T is the overall mean health status in T periods, $w\mathrm{cov}_l$ is the longitudinal weighted covariance, y_i^T is the mean health status of individual i after T periods, and R_i^T is the relative rank of individual i in the distribution of mean incomes after T periods. Jones and López-Nicolás[1] showed that the longitudinal concentration index for the distribution of mean health after T periods can be written as the subtraction of two terms: (i) the weighted sum of cross-sectional concentration indices for each time period (Term 1) minus (ii) the difference between period specific income ranks and ranks for mean income over all time periods and their relationship to health (Term 2):

$$CL = \sum_t w_t CX - \frac{2}{NTY^T} \sum_t \sum_t (y_{it} - y^t)(R_{it} - R_i^T) \ldots\ldots\ldots\ldots (3)$$

Where w_t is within-period mean health status divided by the overall health status in T periods, y_{it} is the health status of individual i at time t, y^t is the within-period mean health status, and R_{it} is the relative rank of individual i in the distribution of incomes at time period t.

Any differences between cross-sectional and longitudinal concentration indices can be measured by the health-related mobility index:

$$M^T = 1 - \frac{CL}{\sum_t w_t CX} \ldots\ldots\ldots\ldots (4)$$

The mobility index itself, M^T, is defined as one minus the longitudinal concentration index (after T periods) divided by the weighted sum of the cross-sectional concentration indices. Therefore, if:

- $M^T > 0$, then the weighted sum of the cross-sectional concentration indices *overestimates* the degree of long-run inequality; or
- $M^T < 0$, then the weighted sum of the cross-sectional concentration indices *underestimates* the degree of long-run inequality.

<u>Decomposing inequality</u>

From a policy perspective, it is insightful to decompose (or unpack) the underlying causes of these inequalities. The cross-sectional and longitudinal concentration indices can be broken down into their explanatory attributes without inferring the direction of causality. To decompose cross-sectional concentration indices, a linear regression model can be used:

$$y_i = \alpha + \sum_{k=1}^{K} \beta_x x_{ik} + \varepsilon_i \ldots\ldots\ldots\ldots (5)$$

where y_i is the health status of individual i, x_{ik} are the determinants of health (e.g., age, education, income, and so on), β_k are the coefficients and ε_i is the error term. The cross-sectional concentration index can then be written as:

$$CX = \sum_k \left(\frac{B_k \bar{x}_k}{\mu} \right) CX_k + \frac{GCX_\varepsilon}{\mu} \ldots\ldots\ldots\ldots (6)$$

where $GCX_{\varepsilon} = \frac{2}{n}\sum_{i=1}^{n}\varepsilon_i R_i$; and

μ is the mean of y, $\bar{x}$ is the mean of x_k, CX_k is the concentration index for x_k, and GCX_{ε} is the concentration index for ε_i. The first term in Equation (6) shows that CX is equal to the weighted sum of concentration indices for the k regressors. The second term – GCX_{ε}/μ – is the inequality in health status that cannot be explained.

For longitudinal data, Jones and López-Nicolás[1] show how the mobility index {Equation (4)} can be decomposed into the contributions of explanatory variables using OLS regression. The mobility index can be written as:

$$M^T = \sum_{k=1}^{K}\hat{\beta}_k \frac{\sum_t \bar{x}_k{}^t CX^t{}_{x_k}}{\sum_t y^t CX_t} M^T{}_{x_k} + residual \ldots\ldots\ldots (7)$$

where $M^T{}_{x_k}$ is the x_k related income mobility index after T periods. The term by which the mobility index of the x_kth regressor after T periods is multiplied takes the form of an *elasticity* of health status with respect to x_k evaluated as the income-related inequality means of x_k and health status. Jones and López-Nicolás refer to this term as the 'inequality-weighted elasticity'.[1] As shown in Equation (7), the mobility index consists of two parts: an explained component and an unexplained or residual component. This decomposition can also be performed using a fixed effects regression model to control for unobserved heterogeneity.

3. HILDA data, imputation, and variable definitions

Sample description

The source of data for this study was responses from 17,375 persons (aged 15 and over) who participated in the longitudinal Household Income and Labour Dynamics Australia (HILDA) Survey between 2001 and 2005 and was included in the General Release 5.1 unit record file made available to researchers. An advantage of using HILDA is that it is one of the largest longitudinal surveys in Australia to have detailed information on income and responses to the SF-36 questionnaire over a five-year period. The HILDA Survey also collects extensive information on household dynamics, labour force participation, employment status, socio-demographics, and well-being. The Survey began in 2001 with a national probability sample of 7,682 responding private households. New waves of data are collected each year, with members of the households responding in each previous wave forming the basis of the panel to be interviewed in the subsequent wave. The HILDA Survey uses several survey instruments, including a household questionnaire, person questionnaires and a self-completion questionnaire. The self-completion questionnaires are collected by the interviewer at a date after the personal interview or, failing that, are returned by mail.

Of the 17,375 persons who participated in HILDA between 2001 and 2005, a total of 9,311 respondents took part in all five waves. We removed 102 respondents who had missing data for any of the household or person questionnaire variables used in this study, or did not have any sample weights assigned to them because they lived in a very remote geographical area. This left a balanced file of 9,209 respondents.

Imputation

The balanced file of 9,209 respondents contained a relatively high degree of missing data for items in the self-completion questionnaire, namely; the health status variable (11.3 per cent missing), the smoking status variable (7.0 per cent missing), and the self-rated prosperity variables (6.7 per cent missing). To account for these missing data we employed a regression-based single imputation technique using the explanatory variables reported in Table 1.

Variable definitions

We employed an algorithm developed by Brazier and colleagues for calculating utility values using the SF-6D, which is derived from the SF-36 into six dimensions.[11] The algorithm is based on a regression equation estimated using a sample of 249 health states that were validated using the standard gamble approach by 611 persons from the general United Kingdom population in face-to-face interviews. Using this approach, a utility score was calculated for each person in each year in the HILDA sample.

Household equivalised disposable income was calculated using the modified Organisation of Economic Co-operation and Development equivalence scale of:

$$household\ equivalent\ income = \frac{household\ income}{1 + 0.5 * (number\ of\ adults - 1) + 0.3 * number\ of\ children}$$

In the above equation, all persons aged 15 or more years were deemed to be adults and all persons aged less than 15 years were deemed to be children.

Household equivalised disposable incomes were then converted to real household equivalised disposable incomes in 2000–01 dollars using the Consumer Price Index. Table 1 provides the names, definitions, and descriptive statistics of the variables used in our analysis. Men and women were analysed separately, and all results have been weighted.

4. Results

Summary statistics

Table 1 reports the summary values for the characteristics of the 9,209 individuals (46.3 per cent males and 53.7 per cent females) in the sample. The mean value for the SF-6D utility index for men (women) was 0.77 (0.75). The standard deviations for men and women were 0.12.

SF-6D concentration and mobility indices

Table 2 reports the concentration and mobility indices for the SF-6D. The results of the cross-sectional concentration index, *CX*, calculated using Equation (3), increased over time from 0.0156 to 0.0240 for males, and from 0.0163 to 0.0228 for females. These results indicate that the degree of cross-sectional pro-rich health inequality increased over time. As a consequence of the increasing *CX* results, the weighted sum of the *CX* results for each time period (Term 1) increases over time for both males and females.

Term 2, viewed in Table 2, is consistently negative for both sexes after the first period, reaching -0.0019 and -0.0017 for males and females respectively in 2005,

indicating that the weighted *CX* results are smaller than the *CL* results. These negative results show that the *CX* results do not fully capture the dynamics of health and its relationship to income because they are unable to show that individuals who are upwardly income mobile are, on average, healthier than individuals who are downwardly income mobile. Correspondingly, the mobility indices for both sexes are negative and increase in absolute value over time to -0.0979 and -0.0884 for males and females respectively in 2005. That the mobility indices increase over time is likely to reflect that the longer the period of time covered by the longitudinal data, the greater the degree that the *CL* results are able to capture long-term dynamics.

Pooled OLS regressions

The pooled OLS regressions results for men and women are reported in Table 3. The Huber-White robust standard errors were adjusted for clustering within-individuals due to the use of longitudinal data. The regression results show that health is positively associated with income, education, marriage and prosperity, and negatively associated with age, retirement, and smoking.

Fixed effects regressions

The fixed effects regressions for men and women are presented in Table 4. Our fixed effects model incorporates age (measured by the square of age) and dummy variables for each year (with 2001 as the reference group) to capture the time trend. The other variables included in the fixed effects model are the same as those included in the OLS model except for the time-invariant regressor, ‘English as first language’, which drops out of the analysis.

For men and women, the results show that there is a statistically significant positive association between health and income, prosperity, and marriage. There is a statistically significant positive time trend for women (all time dummy variables have *p*-values < 0.01). For men, the relationship is not as strong. There is also a statistically significant negative association between health and age squared for both men and women.

Decomposition using OLS regressions

Tables 5 and 6 report the contribution of each regressor to the health-related mobility index for males and females, respectively. The *CL* columns show whether a regressor has either a pro-poor or a pro-rich distribution over the long run. For example, age has a *CL* of -0.0467 for males, which means that age has a pro-poor distribution for males (older males, on average, are lower down the income distribution than younger males). In contrast, education has a *CL* of 0.3614 for females, which means that education has a pro-rich distribution for females.

The *mobility (x)* columns show negative mobility indices whenever the *CL* for the regressor is greater than the weighted *CX*. A negative mobility index indicates that the degree of income inequality related to the regressor increases when we average each person's responses over a longer period of time. The only regressor that has positive mobility indices is income, which has a mobility index of 0.1813 for males and 0.1789 for females.

The relatively large mobility index of -6.9008 for the never married variable for females superficially appears to be of interest, however, on closer inspection the result

is simply caused by the weighted *CX* for this variable being a small number. The weighted *CX* for this variable is only 0.0032 (very slightly pro-rich) – the large mobility index simply reflects that the *CL* is almost eight times larger than the weighted *CX*, but still a modest 0.0257.

The *elasticity (x)* columns show the elasticities that convert the mobility indices for each regressor into contributions to the health-related mobility index. For a regressor to contribute to the health-related mobility index not only does it have to be unequally distributed across the income distribution, it also needs to be correlated with the SF-6D. The self-reported prosperity variable has the largest elasticity of all the regressors (0.3108 and 0.2685 for males and females, respectively) and consequently makes a large contribution to the health-related mobility index for both sexes, despite having *CL* results that, in absolute terms, are smaller than that of several other regressors. Therefore self-rated prosperity contributes strongly to the health-related mobility index not so much because it has a pro-rich distribution, but because it is strongly correlated with health.

Other large contributors to the mobility index for both sexes are the income, age and retired regressors. Income makes a positive contribution to the mobility index of 0.0384 and 0.0363 for males and females respectively, although this positive contribution actually detracts from the mobility index because the index has a negative value. Age has negative contributions to the mobility index of -0.0212 and -0.0295 for males and females respectively, while retirement makes negative contributions of -0.0075 and -0.0054. The negative contributions from the age and retired regressors add to the negative mobility index.

Decomposition using fixed effects regressions

Tables 7 and 8, report the contribution of each regressor to the health-related mobility index for males and females, respectively. The *CL* and *mobility (x)* columns are identical to those listed in the OLS decomposition Tables 5 and 6; although, as necessary part of running a fixed effects regression, the age squared variable has replaced the age variable and consequently a new set of *CL* and *mobility (x)* values appears across the top row of Tables 7 and 8.

The *elasticity (x)* column for men shows that, in absolute terms, the following variables have the largest elasticities: self-reported prosperity (0.1130), the log of income (0.0737), higher education (0.0631), major city (0.0213), married/de facto (0.0194), and age squared (0.0188). For women, the following variables, in absolute terms, have the largest elasticities: self-reported prosperity (0.1000), the log of income (0.0850), age squared (0.0530), higher education (0.0528), married/de factor (0.0403), and widowed (0.0352).

The largest negative *contributors* to the health-related mobility index for men are: higher education (-0.0081), self-reported prosperity (-0.0037), marriage (-0.0025), and age squared (-0.0020). For women, the largest negative *contributions* are: age squared (-0.0069), higher education (-0.0066), self-reported prosperity (-0.0049), and never married (-0.0040). As the health-related mobility index is a negative value, these negative contributions make up a positive share of the mobility index.

5. Discussion

OLS decomposition

The results from this study suggest that individuals who are upwardly income mobile are, on average, healthier than individuals who are downwardly income mobile. However, is this because of their health status or is it because health status is correlated with other variables that are associated with income mobility? Our illustrative OLS decomposition suggests that, for men and women, at least 15.18 per cent and 17.00 per cent of health-related income mobility occurs because health status is correlated with other variables that are associated with income mobility. For males, the variables that contribute most—in order of contribution are—age, self-rated prosperity, retirement status, English as first language status, and educational status. For females, the order of contribution is only slightly different—age, self-rated prosperity, retirement status, educational status and English as first language status. This suggests that healthy individuals are more likely to be upwardly income mobile partly because they tend to be younger, more prosperous, not retired, spoke English as their first language, university educated, non-smokers, and married. The observed variables, however, do not account for the entire health-related mobility index.

Fixed effects decomposition

Our illustrative fixed effects decomposition suggests that, for men and women, at least 5.90 per cent and 6.44 per cent occur because health status is correlated with other variables that are associated with income mobility. This result is lower than the OLS decomposition. One reason for this difference is that our fixed effects specification is controlling for the average difference (i.e., the inter-individual difference) in any observable or unobservable explanatory variables. By design, the

fixed effects model reduces the impact of omitted variable bias. We view the development of fixed effects decomposition models as an important area of future work.

For men, the largest contributors to the health-related mobility index in absolute terms are: income, higher education, self-reported prosperity, marital status, and age squared. For women, the largest contributors in absolute terms are: income, age squared, higher education, self-reported prosperity, and never married. Therefore our fixed effects decomposition suggest that, for both sexes, healthier individuals are more likely to be upwardly income mobile partly because they tend to be younger, more prosperous, and university educated.

Comparison with other studies

Overall, our results (while higher) are broadly comparable to the British[1] and Belgian[7] studies. Our health-related mobility indices for men and women were -0.0979 and -0.0884, respectively. In the British study, the mobility indices for men and women (after five waves) were -0.0624 and -0.0781.[1] In the Belgian study, the corresponding mobility index (after five waves) for men and women combined was -0.0733.[7]

Regression towards the mean

The positive mobility indices for income are likely to be artefacts of the regression towards the mean (RTM) phenomenon that is encountered in longitudinal analysis when analysing variables that are prone to measurement error, such as income. RTM is a statistical phenomenon that can make natural variation in repeated data look like

real change.[18] It happens when unusually large or small measurements tend to be followed by measurements that are closer to the mean. The negative slope of the regression line in Figure 2 shows that regression to the mean is a problem with the income data, because it indicates that respondents that reported a low real income in 2001 had a strong tendency to report a higher income in 2005, while respondents that reported a high real income in 2001 had a strong tendency to report a lower income in 2005.

If actual income (as opposed to reported income) was available for analysis, it is unlikely that it would produce a positive mobility index. The RTM phenomenon is likely to have reduced the share of the health-related mobility index contributed by observed variables because of its distorting impact on the income variable. It seems reasonable to expect, if actual rather than reported income data were available, the mobility index for income would have been negative rather than positive. If the mobility index for income had been negative, the share of the health-related mobility index contributed by the observed variables would have been higher. Furthermore, the income measurement errors would have also had the effect of reducing the observed health-related mobility index, as income mobility would be underestimated.

Although the RTM issue potentially complicates the interpretation of our decomposition analysis, we view the use of panel data in generating longitudinal concentration indices as being critical because it paints a more accurate (and dynamic) picture of the income-related health inequalities. This is primarily because the longitudinal formula essentially averages out individual health status and income responses over time.

Table 1: Names, definitions and descriptive statistics of variables

Variable	Definitions		Males (mean)	Females (mean)
Age	Age last birthday at 30 June of current wave		44.54	45.81
Age^2	The square of Age		2282.81	2414.83
Income (log)	Logarithm of equivalised real household disposable income		4.38	4.35
Education	1 if has university qualification, 0 otherwise		0.17	0.18
Major city	1 if resides in major city, 0 otherwise		0.67	0.67
English first language	1 if English is first language learnt, 0 otherwise		0.87	0.85
Unemployed	1 if unemployed, 0 otherwise		0.04	0.03
Retired	1 if retired, 0 otherwise		0.15	0.17
Married	1 if married or in a de facto relationship, 0 otherwise		0.65	0.63
Divorced	1 if divorced, 0 otherwise		0.04	0.07
Never married	1 if never married and not in a de facto relationship, 0 otherwise		0.26	0.18
Widowed	1 if widowed, 0 otherwise		0.02	0.09
Smoker	1 if smoker, 0 otherwise		0.25	0.20
Prosperity	1 if self-rated prosperity is 'reasonably comfortable' or higher, 0 otherwise		0.67	0.66
SF-6D	Health state utility	Mean	0.77	0.75
		SD	0.12	0.12
		Minimum	0.30	0.30
		Maximum	1.0	1.0
		25th percentile	0.70	0.67
		50th percentile	0.79	0.76
		75th percentile	0.85	0.85
N			4,384	4,951
N*T			21,920	24,755

Notes

- Population weighted results.
- The explanatory variables used in the single imputation regressions were age, income (log), education, major city, English first language, unemployed, retired, married, divorced, widowed.

Table 2: Concentration and mobility indices for SF-6D health state classification

Year	CX	Term 1	Term 2	CL	Mobility index
Males					
2001	0.0156	0.0156	0.0000	0.0156	0.0000
2002	0.0171	0.0164	-0.0005	0.0169	-0.0335
2003	0.0169	0.0166	-0.0011	0.0176	-0.0634
2004	0.0213	0.0177	-0.0014	0.0191	-0.0770
2005	0.0240	0.0190	-0.0019	0.0208	-0.0979
Females					
2001	0.0163	0.0163	0.0000	0.0163	0.0000
2002	0.0187	0.0175	-0.0005	0.0180	-0.0260
2003	0.0191	0.0180	-0.0010	0.0190	-0.0554
2004	0.0210	0.0188	-0.0013	0.0201	-0.0714
2005	0.0228	0.0196	-0.0017	0.0213	-0.0884

Note: Population weighted results.

Table 3: Pooled OLS regression estimates (β) and robust standard errors (RSE). Dependent variable = SF-6D

	Males - (n = 21,290)		Females - (n = 24,755)	
Variable	**β**	**RSE**	**β**	**RSE**
Age	-0.0015***	0.0001	-0.0012***	0.0001
Income (log)	0.0190***	0.0026	0.0180***	0.0025
Education	0.0079**	0.0035	0.0059**	0.0029
Major city	0.0046	0.0030	-0.0065**	0.0027
English first language	0.0213***	0.0050	0.0221***	0.0039
Unemployed	-0.0074	0.0069	-0.0262***	0.0065
Retired	-0.0232***	0.0046	-0.0141***	0.0037
Married	0.0272***	0.0077	0.0209***	0.0070
Divorced	0.0105	0.0095	0.0060	0.0081
Never married	0.0066	0.0084	0.0026	0.0076
Widowed	0.0262**	0.0119	0.0114	0.0084
Smoker	-0.0194***	0.0036	-0.0176***	0.0035
Prosperity	0.0545***	0.0030	0.0489***	0.0027
CONS	0.6838***	0.0144	0.6734***	0.0129
Adjusted R^2	0.1450		0.1154	

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.
Note: Population weighted results.

Table 4: Fixed effects regression estimates (β). Dependent variable = SF-6D

	Males - (n = 21,290)		Females - (n = 24,755)	
Variable	**β**	**RSE**	**β**	**RSE**
Age squared	-0.00003***	0.00001	-0.00007***	0.00001
Income (log)	0.00662***	0.00180	0.00755***	0.00172
Education	0.01779**	0.00760	0.01345*	0.00744
Major city	0.00766*	0.00474	0.00263	0.00459
English first language	(dropped)		(dropped)	
Unemployed	-0.00069	0.00356	-0.01039***	0.00367
Retired	-0.00406	0.00337	-0.00265	0.00263
Married	0.02176***	0.00599	0.00949*	0.00538
Divorced	0.02187***	0.00756	0.00613	0.00627
Never married	0.00773	0.00706	0.01470**	0.00660
Widowed	0.02260*	0.01176	0.01311	0.00821
Smoker	-0.00799***	0.00274	-0.00428	0.00292
Prosperity	0.01982***	0.00183	0.01823***	0.00176
2002	0.00588***	0.00188	0.01076***	0.00181
2003	0.00369	0.00246	0.01257***	0.00238
2004	0.00903***	0.00323	0.01734***	0.00315
2005	0.00694*	0.00409	0.02293***	0.00399
CONS	0.76998***	0.02440	0.84771***	0.02436

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$.
Notes:
Population weighted results.
The reference category for the year dummy variables is 2001.

Table 5: OLS decomposition of the mobility index by factors, males

Variable	CL	Mobility(x)	Elasticity(x)	Contribution (x)	%
Age	-0.0467	-0.1094	0.1942	-0.0212	21.70
Income (log)	0.0303	0.1813	0.2117	0.0384	-39.22
Education	0.3365	-0.1290	0.0282	-0.0036	3.72
Major city	0.0656	-0.0882	0.0128	-0.0011	1.15
English first language	0.0292	-0.1356	0.0327	-0.0044	4.53
Unemployed	-0.3138	-0.0789	0.0053	-0.0004	0.43
Retired	-0.4183	-0.0794	0.0947	-0.0075	7.68
Married	0.0227	-0.1302	0.0243	-0.0032	3.23
Divorced	-0.1696	-0.0784	-0.0049	0.0004	-0.39
Never married	0.0130	-0.0554	0.0015	-0.0001	0.08
Widowed	-0.4504	-0.0814	-0.0155	0.0013	-1.29
Smoker	-0.0595	-0.1909	0.0168	-0.0032	3.28
Prosperity	0.1276	-0.0324	0.3108	-0.0101	10.27
Sum				-0.0149	15.18
Residual				-0.0830	84.82
Total index				-0.0979	100.00

Note: Population weighted results.

Table 6: OLS decomposition of the mobility index by factors, females

Variable	CL	Mobility(x)	Elasticity(x)	Contribution (x)	%
Age	-0.0635	-0.1384	0.2129	-0.0295	33.34
Income (log)	0.0312	0.1789	0.2029	0.0363	-41.07
Education	0.3614	-0.1242	0.0232	-0.0029	3.26
Major city	0.0612	-0.0981	-0.0166	0.0016	-1.84
English first language	0.0210	-0.0990	0.0245	-0.0024	2.74
Unemployed	-0.2654	-0.0589	0.0130	-0.0008	0.87
Retired	-0.4215	-0.0872	0.0619	-0.0054	6.10
Married	0.1011	-0.0157	0.0887	-0.0014	1.58
Divorced	-0.2141	-0.1144	-0.0053	0.0006	-0.69
Never married	0.0257	-6.9008	0.0001	-0.0007	0.82
Widowed	-0.4803	-0.1040	-0.0307	0.0032	-3.61
Smoker	-0.0486	-0.0604	0.0107	-0.0006	0.73
Prosperity	0.1272	-0.0486	0.2685	-0.0130	14.75
Sum				-0.0150	17.00
Residual				-0.0734	83.00
Total index				-0.0884	100.00

Note: Population weighted results.

Table 7: Fixed effects decomposition of the mobility index by factors, males

Variable	CL	Mobility(x)	Elasticity(x)	Contribution (x)	%
Age squared	-0.1057	-0.1090	0.0188	-0.0020	2.09%
Income (log)	0.0303	0.1813	0.0737	0.0134	-13.65%
Education	0.3365	-0.1290	0.0631	-0.0081	8.32%
Major city	0.0656	-0.0882	0.0213	-0.0019	1.92%
Unemployed	-0.3138	-0.0789	0.0005	0.0000	0.04%
Retired	-0.4183	-0.0794	0.0166	-0.0013	1.35%
Married	0.0227	-0.1302	0.0194	-0.0025	2.59%
Divorced	-0.1696	-0.0784	-0.0102	0.0008	-0.82%
Never married	0.0130	-0.0554	0.0017	-0.0001	0.10%
Widowed	-0.4504	-0.0814	-0.0134	0.0011	-1.11%
Smoker	-0.0595	-0.1909	0.0069	-0.0013	1.35%
Prosperity	0.1276	-0.0324	0.1130	-0.0037	3.73%
Sum				-0.0058	5.90%
Residual				-0.0921	94.10%
Total index				-0.0979	100.00%

Note: Population weighted results

Table 8: Fixed effects decomposition of the mobility index by factors, females

Variable	CL	Mobility(x)	Elasticity(x)	Contribution (x)	%
Age squared	-0.1376	-0.1310	0.0530	-0.0069	7.86%
Income (log)	0.0312	0.1789	0.0850	0.0152	-17.20%
Education	0.3614	-0.1242	0.0528	-0.0066	7.42%
Major city	0.0612	-0.0981	0.0067	-0.0007	0.74%
Unemployed	-0.2654	-0.0589	0.0052	-0.0003	0.34%
Retired	-0.4215	-0.0872	0.0116	-0.0010	1.15%
Married	0.1011	-0.0157	0.0403	-0.0006	0.72%
Divorced	-0.2141	-0.1144	-0.0055	0.0006	-0.71%
Never married	0.0257	-6.9008	0.0006	-0.0040	4.58%
Widowed	-0.4803	-0.1040	-0.0352	0.0037	-4.14%
Smoker	-0.0486	-0.0604	0.0026	-0.0002	0.18%
Prosperity	0.1272	-0.0486	0.1000	-0.0049	5.50%
Sum				-0.0057	6.44%
Residual				-0.0827	93.56%
Total index				-0.0884	100.00%

Note: Population weighted results

Figure 1: Concentration Curve

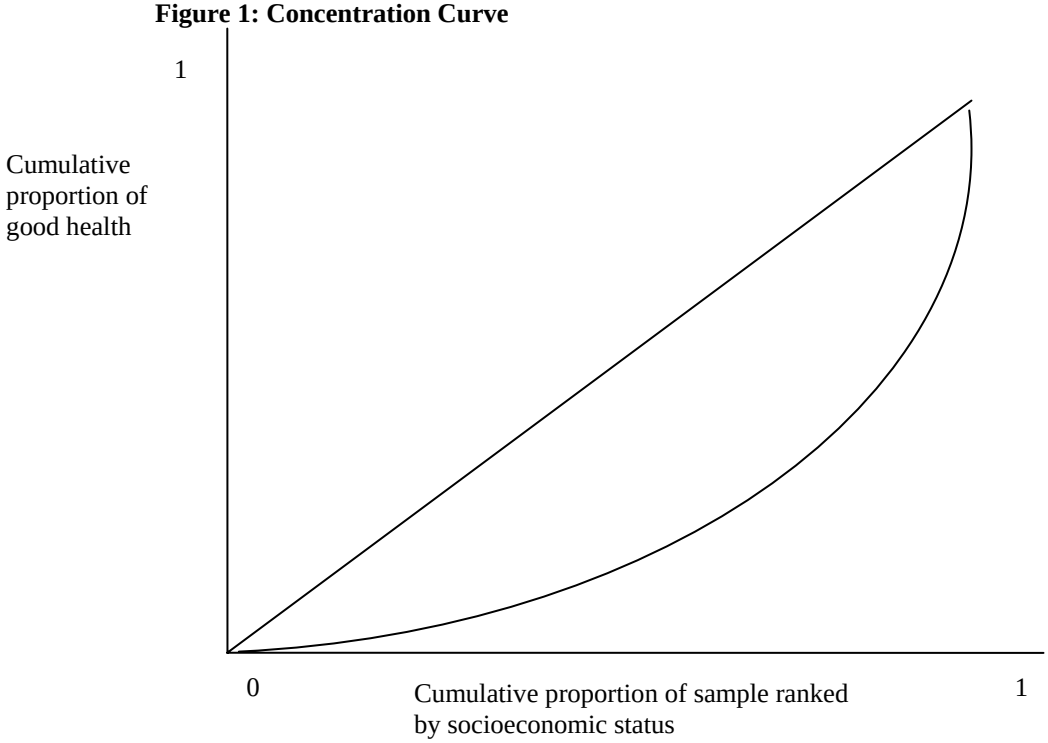
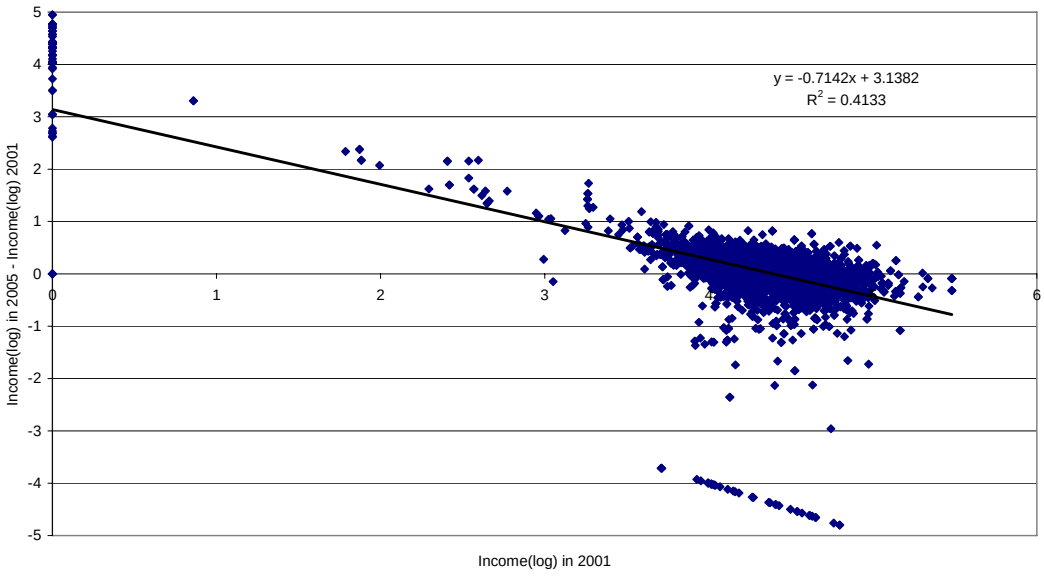


Figure 2: Regression towards the mean: Income (log) in 2000/01 dollars



Note: Population weighted results

References

1. Jones, A.M. and A. López-Nicolás, *Measurement and explanation of socioeconomic inequality in health with longitudinal data.* Journal of Health Economics, 2004. **13**: p. 1015–1030.
2. Wagstaff, A., P. Paci, and E. van Doorslaer, *On the measurement of inequalities in health.* Social Science & Medicine, 1991. **33**: p. 545–557.
3. Wagstaff, A. and E. van Doorslaer, *Measuring inequalities in health in the presence of multiple-category morbidity indicators.* Health Economics, 1994. **3**: p. 281–291.
4. van Doorslaer, E., et al., *Income-related inequalities in health: some international comparisons.* Journal of Health Economics, 1997. **16**: p. 93–112.
5. Kawani, N.C., A. Wagstaff, and E. van Doorslaer, *Socioeconomic inequalities in health: measurement, computation and statistical inference.* Journal of Econometrics, 1994. **77**: p. 87–104.
6. Wagstaff, A., E. van Doorslaer, and N. Watanabe, *On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam.* Journal of Econometrics, 2003. **112**: p. 207–233.
7. Lecluyse, A., *Income-related health inequality in Belgium 1994-2001*, in *ECuity III Working Paper*. 2004.
8. van Doorslaer, E. and A.M. Jones, *Inequalities in self-reported health: Validation of a new approach to measurement.* Journal of Health Economics, 2003. **22**: p. 61–87.

9. Idler, E.L. and Benyamini, *Self-rated health and mortality: a review of twenty-seven community studies.* Journal of Health and Social Behaviour, 1997. **38**: p. 21–37.

10. Ware, J.E., et al., *SF-36 Health Survey Manual and Interpretation Guide*. 1993, Boston: The Health Institute.

11. Brazier, J., J. Roberts, and M. Deverill, *The estimation of a preference-based measure of health from the SF-36.* Journal of Health Economics, 2002. **21**: p. 271–292.

12. Hollingworth, W., et al., *The practicality and validity of directly elicited and SF-36 derived health state preferences in patients with low back pain.* Health Economics, 2002. **11**: p. 71–85.

13. Lundberg, L., et al., *The relationship between health-state utilities and the SF-12 in a general population.* Medical Decision Making, 1999. **19**: p. 128–140.

14. Franks, P., et al., *Mapping the SF-12 to preference-based instruments: convergent validity in a low-income, minority population.* Medical Care, 2003. **41**: p. 1277–1283.

15. Fryback, D.G., et al., *The Beaver Dam health outcomes study: initial catalogue of health-state quality factors.* Medical Decision Making, 1993. **13**: p. 89–102.

16. Nichol, M.B., N. Sengupta, and D.R. Globe, *Evaluating quality adjusted life years: estimation of the Health Utility Index (HUI) from the SF-36.* Medical Decision Making, 2001. **21**: p. 105–112.

17. Sengupta, N., et al., *Mapping the SF-12 to the HUI and VAS in a managed care population.* Medical Care, 2004. **42**: p. 927–937.

18. Barnett, A., J.C. van der Pols, and A.J. Dobson, *Regression to the mean: what it is and how to deal with it.* International Journal of Epidemiology, 2005. **34**: p. 215–220.