Longitudinal analysis of income-related health inequalities: methods, challenges and applications

Martin Siegel  Paul Allanson

This is the pre-peer reviewed version of the following article: Siegel, M. and Allanson, P. (2015) Longitudinal analysis of income-related health inequalities: methods, challenges and applications. Expert Review of Pharmacoeconomics & Outcomes Research, 16 (1), 41-49, which has been published in final form at DOI: 10.1586/14737167.2016.1123096.

Abstract  Socioeconomic inequalities in health are an important research area in health economics and public health. The concentration index has become a well-established measure of income-related health inequalities, and a number of approaches to identify potential causes of health inequalities exist. With the increasing availability of suitable longitudinal data, more sophisticated approaches to monitor inequalities and to identify potential causal relationships between socioeconomic status and health evolved. We first review the concentration index and some more basic approaches to explain health inequalities. We then discuss advantages and potential shortcomings of static and dynamic health inequality measures. We review different concepts of health and socioeconomic mobility, as well as recent studies on the life course perspective and economic changes. Our aim is to provide an overview of the concepts and empirical methodologies in the current literature, and to guide interested researchers in their choice of an appropriate inequality measure.

Keywords health inequalities; concentration index; health-related income mobility; income-related health mobility; longitudinal analyses; life course perspective; economic development
Introduction

Analyzing health inequalities has become an important task for health economists and public health researchers over recent decades. An ever-growing body of literature provides evidence that lower socioeconomic status is commonly associated with poorer health. The growing number of countries with longitudinal datasets comprising socioeconomic and health related information has stimulated the development and refinement of different approaches to the measurement of health inequalities. Researchers from the fields of health economics, public health and related disciplines increasingly address empirical questions concerning, for example, the impact of economic and social political developments on health inequalities, the trajectories of health and socioeconomic status over the life course, and the direction of potential causal relationships between socioeconomic status and health. In this review, we first introduce the concentration index as a common measure of income-related health inequalities and its refinements. We then move on to examine the range of conceptually distinct ‘mobility’ measures that have been put forward to explore various aspects of the distributional consequences of the dynamic inter-relationship between health and socioeconomic status. Finally, we provide a brief overview of empirical applications of different longitudinal measures of health inequalities, focusing on two of the main substantive issues in the literature: life course perspective analyses and the evolution of income-related health inequalities in the light of economic and social changes over time.

Static measures of income-related health inequalities

The concentration index

A common approach to measure income-related inequalities in the field of health economics is the concentration index [1–3]. It was derived from the well-known Gini index: the difference is that the concentration index does not require the outcome variable for which one measures inequality to also be used for the ranking of individuals. The concentration index is based on the concentration curve which was derived from the Lorenz curve and compares the cumulative share of some health
outcome against the cumulative share of the population ranked by socioeconomic status. The concentration curve lies below (above) the 45 degree line of equality if the health variable is concentrated among the better-off (worse-off). Figure 1 provides a hypothetical concentration curve where the health variable is concentrated among the poor as may be the case with illness and disability. In this example, the curve indicates that 30% of the health outcome are concentrated among the poorest 20% of the population, while only 10% of the health outcome are concentrated among the richest 20% of the population (note that this means a three times higher prevalence among the poorest 20% than among the richest 20%).

Figure 1: Hypothetical concentration curve (source: the authors)

The concentration index is a rank-dependent measure of inequality, which reflects the socioeconomic dimension to inequalities in health because individuals are ranked by their income or socioeconomic status (not by their health). It measures twice the area between the concentration curve and the 45 degree line of perfect equality; i.e. how far the observed income-related distribution of health deviates from a scenario of perfect equality. The index is bounded in the (−1; 1) interval and is positive (negative), if the health outcome is concentrated among the
rich (poor). It equals 0 if the concentration curve coincides with the line of perfect equality such that there is no income-related inequality (or if positive and negative areas cancel each other out) [1–5]. The concentration index for the income-related distribution of health shown in Figure 1, for example, would be −0.2, indicating a comparatively strong concentration of the health outcome among the worse-off.

Correcting the concentration index for restricted health variables

A potential problem may arise when the health variable is restricted [6–10]. While the concentration index for an unrestricted variable is asymptotically bounded in the (−1; 1) interval, this does not hold true for restricted health outcomes: Wagstaff [6] has shown that the bounds of the concentration index for binary health outcomes depend inversely on the mean. Thus, comparing health inequalities between populations with rather different overall population health (e.g. the prevalence of a disease or the share of immunized infants) without accounting for the prevalence-dependency of the underlying inequality measure may lead to misleading results [6–10]. Moreover, the choice as to whether one measures health attainments or shortfalls may also have considerable impact on the results [9, 10]. The same holds true for longitudinal analyses of health inequalities in which the level of the health variable may change over time or, in case of several chronic diseases, vary considerably with age [11–13].

Wagstaff [6] and Erreygers [7], among others, propose correction formulas for the concentration index to allow meaningful comparisons of health inequalities across different countries with different overall population health. While all correction methods aim to produce comparable results, all of them have rather different properties and implications [6–10]. The key criterion when deciding how to account for variations in mean health is one’s notion of the most unequal scenario; a rather normative decision [9, 10]. Erreygers [7] provides a technical discussion of a set of alternative approaches, while an intuitive discussion of their various properties, interpretations and policy implications can be found in Kjellsson et al. [14].
Decomposition of the concentration index

When measuring inequalities, one may not only be interested in the extent, but also in the potential causes of income-related health inequalities. One advantage of the concentration index is that it can readily be decomposed into the contributions of factors assumed to influence the health outcome. The basis for this decomposition approach are standard regression models where health is regressed on a set of explanatory variables such as age, sex, income, education, employment status etc. [3, 15, 16]. The estimated contribution of each health determinant is then given as the product of the elasticity of health with respect to the determinant (i.e. the estimated percentage change of health statistically induced by a one percent change in the explanatory variable) and the concentration index of the health determinant with respect to income. The intuitive interpretation is that a (socioeconomic) determinant of health can only contribute to income-related health inequalities if the factor is statistically associated with health and concentrated among the rich or poor. The contributions indicate how the concentration index would differ if the respective explanatory variable was not associated with health or was equally distributed across income groups [17].

The decomposition based on a linear regression model introduced by Wagstaff et al. [15] is a mathematically exact decomposition: the contributions of the explanatory variables and of the error term exactly add up to the observed concentration index of the health variable. As health variables are often restricted, the linear regression based model seems unsuitable in many cases. van Doorslaer et al. [16] suggest using the partial effects of the explanatory variables on health derived from nonlinear regression models in place of the coefficients from linear regression models to allow the decomposition of concentration indices for restricted health variables. A formal introduction to the linear regression based decomposition approach can be found in Wagstaff et al. [15], with the textbook by O’Donnell et al. [3] providing a detailed and fairly intuitive introduction to both linear and nonlinear approaches.
Accounting for heterogeneity in the decomposition of health inequalities

Given data availability, one may also undertake decompositions of the concentration index based on more sophisticated regression models. For example, Jones and López Nicolás [18] further decompose the error term from the above described decomposition of the concentration index [15] to extract inequalities induced by group-specific heterogeneity. Using a fixed effects model with group-specific random slope coefficients, they incorporate income-related variations in the estimated slope coefficients to further decompose the formerly unexplained part (residual term) from Wagstaff et al.'s [15] model. This approach can be considered to be somewhere between cross-sectional and longitudinal methods: the model may be used to exploit the additional information included in longitudinal data, but Jones and López Nicolás [18] have shown that it may also be applied to cross-sectional data (they pool longitudinal data covering several years into one larger cross-sectional dataset and allow the coefficients to vary between pre-defined age groups).

Analysis of income-related health inequalities over time

Static versus dynamic approaches

The standard approach to the analysis of changes in income-related health inequality has been to first use the concentration index and its decomposition to identify the determinants of health inequalities at a number of specified points in time and then to analyze these repeated cross-sectional results to identify the possible causes of change over time (see e.g. Wagstaff et al [15]; Gravelle and Sutton [19]; O’Donnell et al. [3]). This ‘static’ approach has some advantages: the methods are easily comprehensible and well-established among health economists, data requirements are comparably low since it only requires repeated cross-sectional data, and the results are easily communicable. Indeed, even if longitudinal or panel data is available then the static approach may be sufficient to monitor changes in health inequalities over time if one is only interested in the development of groups rather than individuals and the specific composition of those groups does not matter [17]. Nevertheless one should keep in mind that it may involve some shortcomings as the use of cross-sectional data does not allow one to track
individual experiences over time. As a consequence, causal inference on the dynamic inter-relationship between socioecoeconomic status and health is impossible, and one cannot distinguish persistent or chronic inequalities related to ongoing socioeconomic and health-related deprivation from transitory inequalities arising from temporary episodes of poor health and poverty [20].

The inherent limitations of the static approach have led to the emergence of a small but growing literature that employs longitudinal data to characterize and measure the dynamics of socioeconomic-related health inequalities. An important feature of this literature is that it explores not one but several different aspects of the distributional consequences of the dynamic inter-relationship between health and socioeconomic status, leading to the definition of a number of conceptually distinct ‘mobility’ measures. In particular, following the classification of income mobility measures proposed in Jantti and Jenkins [21], it is possible to identify four distinct concepts that may be of normative significance: impact on longer-term income-related health inequality, changes in individual health outcomes, positional change within the socioeconomic distribution and health risk. The need for a multiplicity of measures should come as little surprise in the light of the complex, multifaceted nature of the phenomenon.

**Transitory or chronic health inequalities?**

In a pioneering paper on the use of longitudinal data to analyze income-related health inequalities, Jones and López Nicolás [22] explore how the value of the concentration index changes as the measurement period is extended from one to several periods. For this purpose they propose an index of ‘health-related income mobility’, modeled on the Shorrocks [23] income mobility index, which measures the extent to which the concentration index is larger or smaller in the short run than in the long run. Using British Household Panel Survey (BHPS) data for Great Britain they find that the concentration index is typically larger the greater the number of years over which measurements are taken, implying that long-term or chronic problems of income-related health inequalities are typically more severe than would be inferred from short run or cross-sectional estimates. This finding has subsequently been reproduced in a number of other studies using health data sets for a range of developed countries (see e.g.
Hernandez-Quevado et al. [24]; Lecluyse [25]; Brandrup and Kortt [26]. Allanson et al. [20] demonstrate that the empirical regularity is likely to be explained by the existence of a stronger positive association between permanent disparities in income and health across individuals than between short-run changes in the income and health of individuals. By implication, policies designed to tackle income-related health inequalities need to address structural problems that trap some individuals in deprivation and ill-health and not just deal with transitory episodes of poverty and sickness.

**Characterizing processes of inequality change**

A second main use of longitudinal data has been to characterize the process of distributional change underlying the evolution of cross-sectional income-related health inequalities over time. The degree of mobility in individual health outcomes per se can be assessed descriptively using inter-temporal correlation and transition matrices [27], with an accompanying range of summary measures available in the literature (see Fields [28], for a discussion). Transition matrices have also been employed to examine the impact of socioeconomic mobility on health inequalities by comparing the health of individuals who change socioeconomic class with that of those who remain in the same class. For example, Boyle et al. [29] show that health inequalities between classes may widen even if the health of those who change (deprivation) class lies on average somewhere between that of the class they left and the class they joined.

A starting point for understanding such findings is provided by the observation in Allanson et al. [20] that any change in income-related health inequality in a fixed population must arise from some combination of changes in health outcomes and changes in individuals’ positions in the income distribution. By decomposing the change in concentration index between two periods, they provide an index of ‘income-related health mobility’ that captures the effect of differences in relative health changes between individuals whose initial income differs. The index addresses the question of whether the pattern of health changes favor those with initially high or low incomes, providing a natural counterpart to the concentration index that address the issue of whether those with better health tend to have high or low incomes. They also obtain an index of ‘health-related income mobility’, which is different from the index of the same name.
defined in Jones and López Nicolás (2004) in that it captures the effect of the reshuffling of individuals within the income distribution on cross-sectional health inequalities. This reranking effect is expected to be disequalizing since those who move up the income distribution tend to be healthier in the final period than those who move down.

Allanson and Petrie [30] subsequently establish the normative implications of this type of decomposition by embedding the change in the concentration index within a broader analysis of the change in the health of the population. In particular, evaluating income-related health mobility on the basis of the social weights associated with individuals’ initial income ranks is shown to give greater weight to the health prospects of those that start with lower income, where this asymmetric treatment of individuals may be justified on the grounds that the initially poor are disadvantaged to the extent that they face a worse lottery of future health possibilities than those who are better off. The paper further demonstrates that the same decomposition procedure can also be used to analyze changes in other rank-dependent income-related health inequality indices, and with inequalities measured with respect to health shortfalls rather than attainments. Allanson and Petrie [10] illustrate how the choice of health inequality measure implies a particular vertical equity judgment, which may be expressed in terms of a health inequality equivalence criterion that specifies how a given change in population health should be distributed so as to leave health inequality unchanged from its initial value (see Kjellsson and Gerdtham [31]; Kjellsson et al. [14]; for further discussion).

The original decomposition in Allanson et al. [20] is extended to account for mortality in Petrie et al. [32] and for all other sources of population change, including births and migration, in Allanson and Petrie [10]. The latter study employs BHPS to analyze the source of differences in cross-sectional income-related health inequalities in Great Britain between 1999 and 2004. It is found that an increase in health inequalities among the adult population resident over the study period was partially masked by the impact of both selective mortality, as the poor and sick were more likely to die, and by the entry of youths into the adult population. The need to take the effects of migration into account is also demonstrated even though immigrant population groups in particular are typically under-represented in longitudinal household surveys such as the BHPS. The conclusion is reached that failure to take the confounding effects of demographic
and population change into account may lead to erroneous conclusions on the effectiveness of policies to tackle health inequalities.

The decomposition of changes in the concentration index serves to characterize the underlying mobility processes and thereby provides a first step to understanding the dynamics of income-related health inequalities. Allanson and Petrie [33] go on to investigate the role of individual health determinants in driving changes in health inequality through both morbidity changes and mortality. Specifically, BHPS data for the period 1999 to 2004 is employed to estimate a Two-Part Model – a probit model of survival together with a dynamic health function conditional upon survival – that is in turn used to identify the contribution of each health determinant to income-related health mobility. Health changes due to expected mortality were found to account for the bulk of overall income-related health mobility, with the major driver of the disequalizing effects of mortality being the positive association between (old) age and poverty given that the old were at greater risk of death, and with other significant contributors including initial health status, education, gender and smoking. Moreover, morbidity changes also had a disequalizing effect despite the poor enjoying a disproportionate share of contemporaneous health gains due to pro-poor real income growth over the period. These findings point to the importance of understanding the determinants of health changes in the design and evaluation of policies designed to tackle health inequalities, and might usefully be complemented by a parallel decomposition of health-related income mobility based on a model of income changes.

*Mobility as risk*

The final mobility concept is motivated by a concern with predictability rather than movement per se, originating from the insight that greater mobility may no longer be regarded as socially desirable if it is associated with more pronounced intertemporal fluctuations and more uncertainty [23, 34]. In particular, individuals with a preference for health stability and an aversion to risk may choose a future health trajectory with a lower present value if it is both less volatile and more certain. Hauck and Rice [27] propose two health mobility indices, based on analogous measures of income risk, which are both defined such that lower index values are associated with greater transitory
variation in health and hence greater health risk. The first index is based on the estimation of an error components panel data model to partition residual variability in health states into permanent and transitory components, with the measure defined as the proportion of residual variability attributable to the permanent component. The second measure is given by the estimated coefficient on lagged health status from a dynamic panel data model, which is informative about the degree of dependence between previous health and current health status and therefore indicative of the degree of health persistence. Using BHPS data for Great Britain, evidence is found that mobility in mental health differs systematically across socioeconomic groups, with individuals from lower income groups associated with both poorer mental health and lower health mobility. Hauck and Rice [27] stress the adverse implications of health persistence for those among the poor with mental health problems, but it should also be borne in mind that the better off face higher instability and risk due to the greater unpredictability of their mental health outcomes.

Empirical applications

One may consider several potential research questions where a longitudinal measure of income-related health inequalities may be desirable or even crucial. This section illustrates two of the main issues in the empirical literature where the dynamics of socioeconomic inequalities in health were investigated: life course perspective analyses and the evolution of income-related health inequalities in the light of social and economic changes over time. A brief introduction of the theoretical background and the key research questions is followed by a description of how the dynamics of income-related health inequalities were addressed in some of the recent papers from these two strands of the literature.

Life course perspective analyses of health inequalities

The literature on life course perspectives in income-related health inequalities has long been dominated by consideration of the disadvantage accumulation and age as leveler hypotheses, which have often been seen as competing explanations of
why income-related disparities in health may differ over the life course though this is not necessarily the case [35]. Both agree that lower socioeconomic status and lacking resources are associated with less healthy lifestyles, higher health risks, a faster decline of health status and higher mortality rates. The disadvantage accumulation hypothesis contends that social gradients in health develop in early life and become stronger as socioeconomic and health disadvantages accumulate over the complete life course [36–41]. The age as leveler hypothesis is consistent with the disadvantage accumulation hypothesis to some extent, but adds the assumption that the decline of health is an unavoidable part of aging. Health inequalities evolving from socioeconomic disadvantages increase up to some point in midlife, but disparities narrow in older age groups owing to a decline in health status due to aging among the elderly [37, 42–47]. The two mechanisms are not necessarily contradictory [35], and evidence found for the age as leveler hypothesis may also be consistent with the disadvantage accumulation hypothesis [37]. Note that selective mortality may be an important potential confounder in empirical life course analyses [48]. If the sick and poor die at younger ages and only the rich and healthy remain in the data, this may lead to an artificial leveling of health inequalities among the elderly.

In the context of this debate, Siegel and Mosler [12] combine the concentration index with nonparametric regression techniques to measure variations in income-related health inequalities over the life course. Applying this model to self-assessed health [12], obesity, hypertension and diabetes [11] yields that the concentration of ill-health among the economically worse-off increases until mid-life and decreases in later life. The almost monotonously increasing prevalence in all measures of ill-health with age supports the notion that the prevalence of ill-health rises first among the economically deprived until the resulting income-related health inequalities are leveled out in later life when ill-health also increases among the better-off [11, 12, 37, 48]. The comparably low mortality rates in those age groups where disparities decline suggest that the observed patterns are unlikely to be solely an artificial result of selective mortality [11, 12].
**Income-related health inequalities in the light of economic and social change**

Another prominent strand in the empirical literature on longitudinal analyses of income-related health inequalities addresses developments of social, political and economic contexts over time. Nolte and McKee [49], for example, address the impact of the German reunification on health inequalities in East and West Germany. Using data for the years 1992 and 1997, they choose a cross-sectional approach and compare differences in the associations between income and health for both parts of Germany for the two years. Siegel et al. [17] apply the regression-based decomposition approach to repeated cross-sectional data separately for each year between 1994 and 2011 to investigate changes in income-related health inequalities in Germany over time. They argue that the repeated cross-sectional approach assures a representative sample of the working population throughout the study period as it avoids potential problems with aging. Using an earlier version of the same dataset, Kroll and Lampert [50] investigate health inequalities with respect to employment status. They use a more longitudinal approach and estimate a fixed-effects model, where they interpret an interaction effect between time and employment status as a trend in inequalities. To our knowledge, no studies using rank-dependent health inequality measures to address potential effects of the European economic crisis which started in 2008 have been published to date. Although Siegel et al. [17] include its beginning in their study period, potential consequences of the crisis have not been in their research focus.

Addressing the effects of an aging society on income-related health inequalities, Kamrul Islam et al. [51] use Swedish data and a fixed-effects panel data estimator for the decomposition of the concentration index. They conclude that, in aging societies, a decline of mean health may increase income-related health inequalities. They highlight two potential effects that may bias the results: the pension effect may lead to higher health inequalities as the elderly are on average poorer (with pensions being lower than wages) and less healthy. The student effect, on the other hand, may bias the observed inequalities downwards as younger individuals (students and young individuals at the beginning of their working life) are on average poor but healthy.

van Oorti et al. [52] theoretically and empirically investigate potential effects of economic growth on income-related health inequalities. They argue that effects of
changes in mean income and in income inequalities on income-related health inequalities strongly depend on the elasticity of income on health and distinguish direct and indirect effects of income growth and income inequalities on health and health inequalities. The sign of the direct effects can be derived a priori: if income elasticity on health increases on average, income-related health inequalities will grow with income and income inequality. In contrast to the direct effects, however, the indirect effects can not be anticipated from the theoretical model but can only be observed empirically [52]. van Ourti et al. [52] conclude that income growth increases the mean population health as long as income inequality does not increase. At the same time, they find that only decreasing income inequality may decrease income-related health inequalities.

**Expert commentary**

It is crucial to keep in mind that there are important aspects of income-related health inequality changes that are not revealed by examining changes in cross-sectional data over time. For example, if half the population of a country is always poor and sick while half is rich and healthy then it will not be possible to determine whether it is always the same individuals in each category. More generally, it will not be feasible to distinguish between income-related health inequalities arising from chronic or persistent social disadvantage as opposed to transitory episodes of both poverty and sickness, where the former state might be deemed less socially acceptable than the latter. Moreover longitudinal data are required to determine the incidence and effectiveness of interventions designed to tackle such health inequalities in the population, where chronic inequalities might call for policies to tackle the structural problems that trap some individuals in deprivation and ill-health while transitory episodes might demand measures such as improvements in acute health services or temporary welfare assistance.

**Five-year view**

Although a number of the earlier studies on income-related health inequalities used comparatively simple regression models, the literature has evolved and econometric techniques become more sophisticated. Two big issues, however, warrant further
examination: causality and heterogeneity. Firstly, the directions of causal relationships may further be addressed by researchers with access to longer data panels and larger datasets. Panel data models, nonparametric regression and Bayesian estimation techniques may improve the power and accuracy of econometric models and may provide further insight into potential causal pathways generating income-related health inequalities. Secondly, heterogeneities in health inequalities and in the associations between determinants of health and health have long been regarded as a merely technical issue involving potential biases, with researchers developing sophisticated ‘black box’ approaches to account for them. Some part of the future research on health inequalities should explicitly study variations in health inequalities, as well as variations in the contributions of different determinants of health across cohorts, age-groups or different regions. Increasing availability of data and computational power will enable researchers to apply more complex models and techniques; being aware of models and estimation approaches derived in other research areas such as financial econometrics, marketing, psychology, biometrics, psychometrics and epidemiology, may become even more important in the future.

**Key issues**

- Longitudinal analyses of health inequalities can use static and dynamic models.
- Static models (e.g. repeated cross-sections) may be useful to compare groups of the population if changes in the composition of such groups does not matter.
- Static measures do not allow to distinguish between transitory inequalities (short episodes of ill-health and poverty) from ongoing structural socioeconomic and health-related deprivation.
- Dynamic methods are required to explore various distinct aspects of the distributional consequences of the dynamic inter-relationship between health and socioeconomic status.
- In particular, dynamic measures allow one to distinguish between transitory and chronic health inequalities and to characterize processes of inequality change.
• Static and dynamic measures have their advantages and disadvantages, and the right choice depends on the underlying research question.

• Heterogeneity in health inequalities, as well as in the associations between the determinants of health and health, should regarded as more than a potential bias to be controlled for.

References


• A textbook worth reading: it is an extensive, easily accessible introduction to the field of health inequality measurement.


- This paper provides an accessible and practice-orientated treatment of the choice of inequality measure.


- The key paper introducing the decomposition approach.


- To our knowledge, this is currently the only published paper applying a rank-based measure of health inequalities to data covering some years of the economic crisis which started in 2008.


- A key reading for the dynamic measurement of health inequalities, the paper also introduces an alternative characterization of processes underlying health inequality changes.


- This paper applies the concentration index approach to the life course perspective, thereby giving a good discussion of the literature on life course analyses.


