



*Demographic Research* a free, expedited, online journal  
of peer-reviewed research and commentary  
in the population sciences published by the  
Max Planck Institute for Demographic Research  
Konrad-Zuse Str. 1, D-18057 Rostock · GERMANY  
[www.demographic-research.org](http://www.demographic-research.org)

---

, citation and similar papers at [core.ac.uk](http://core.ac.uk)

brought to

provided by Research P

**DEMOGRAPHIC RESEARCH**

VOLUME 8, ARTICLE 11, PAGES 305-358

PUBLISHED 17 JUNE 2003

[www.demographic-research.org/Volumes/Vol8/11/](http://www.demographic-research.org/Volumes/Vol8/11/)

DOI: 10.4054/DemRes.2003.8.11

*Research Article*

**Gini coefficient as a life table function:  
computation from discrete data,  
decomposition of differences and  
empirical examples**

**Vladimir M. Shkolnikov**

**Evgueni E. Andreev**

**Alexander Z. Begun**

© 2003 Max-Planck-Gesellschaft.

## Table of Contents

1	Introduction	306
2	Definitions and properties	308
2.1	The Lorenz curve	308
2.2	Application to the distribution of length of life	308
2.3	Gini coefficient	310
2.4	Basic properties of inequality measures	313
3	Empirical trends in Gini coefficient and other measures of inequality and judgements about direction of changes in inequality	314
4	Computation of Gini coefficient from complete and abridged life tables	318
5	Decomposition of a difference between two values of Gini coefficient	324
5.1	Age	324
5.2	Age and cause of death	328
5.3	Age, mortality and population structure	332
6	Variations in life expectancy and Gini coefficient in time and across countries	338
7	Conclusion	343
8	Acknowledgements	345
	Notes	346
	References	347
	Appendix 1	352
	Appendix 2	353
	Appendix 3	354
	Appendix 4	356

Research article

## **Gini coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples**

**Vladimir M. Shkolnikov**<sup>1</sup>

**Evgueni E. Andreev**<sup>2</sup>

**Alexander Z. Begun**<sup>3</sup>

### **Abstract**

This paper presents a toolkit for measuring and analyzing inter-individual inequality in length of life by Gini coefficient. Gini coefficient and four other inequality measures are defined on the length-of-life distribution. Properties of these measures and their empirical testing on mortality data suggest a possibility for different judgements about the direction of changes in the degree of inequality by using different measures. A new computational procedure for the estimation of Gini coefficient from life tables is developed and tested on about four hundred real life tables. The estimates of Gini coefficient are precise enough even for abridged life tables with the final age group of 85+. New formulae have been developed for the decomposition of differences between Gini coefficients by age and cause of death. A new method for decomposition of age-components into effects of mortality and composition of population by group is developed. Temporal changes in the effects of elimination of causes of death on Gini coefficient are analyzed. Numerous empirical examples show: Lorenz curves for Sweden, Russia and Bangladesh in 1995, proportional changes in Gini coefficient and four other measures of inequality for the USA in 1950-1995 and for Russia in 1959-2000. Further shown are errors of estimates of Gini coefficient when computed from various types of mortality data of France, Japan, Sweden and the USA in 1900-95,

---

<sup>1</sup> Max Planck Institute for Demographic Research. Correspondence to: Max Planck Institute for Demographic Research, Konrad-Zuse-Str., 1, D-18057 Rostock, Germany. Tel: (49 381) 2081 147; Fax: (49 381) 2081 447. E-mail: Shkolnikov@demogr.mpg.de

<sup>2</sup> Centre of Demography and Human Ecology, Moscow. E-mail: evg\_andreev@ns.cnt.ru

<sup>3</sup> Universität der Bundeswehr Hamburg. E-mail: begun@unibw-hamburg.de

decompositions of the USA-UK difference in life expectancies and Gini coefficients by age and cause of death in 1997. As well, effects of elimination of major causes of death in the UK in 1951-96 on Gini coefficient, age-specific effects of mortality and educational composition of the Russian population on changes in life expectancy and Gini coefficient between 1979 and 1989. Illustrated as well are variations in life expectancy and Gini coefficient across 32 countries in 1996-1999 and associated changes in life expectancy and Gini coefficient in Japan, Russia, Spain, the USA, and the UK in 1950-1999. Variations in Gini coefficient, with time and across countries, are driven by historical compression of mortality, but also by varying health and social patterns.

## **1. Introduction**

At present, the average level of length of life is high in many countries and it is interesting to study to what extent this advantage is equally accessible to all people. This is why measures of variability in respect to length of life attract growing attention (Anand et al., 2001).

Gini coefficient is the most common statistical index of diversity or inequality in social sciences (Kendall and Stuart, 1969, Allison, 1978). It is widely used in econometrics as a standard measure of inter-individual or inter-household inequality in income and wealth (Atkinson, 1970 and 1980, Sen, 1973, Anand, 1983). Gini coefficient can also be used as a measure of inequality in length of life (or as a degree of inter-individual variability in age at death).

In a number of studies, Gini coefficient has been applied to mortality schedules. In some studies, Gini coefficient has been used to measure variability in levels of mortality among socio-economic groups (Leclerc et al., 1990). However, in most studies it expressed inter-individual variability in age at death (Le Grand, 1987, 1989, Illsey and Le Grand, 1987, Silber, 1988, 1992, Llorca et al., 1998).

Illsey and Le Grand (1987), who justified the use of Gini coefficient for the analysis of inequality in health in the 1980s, stressed that the individual-based measurement of inequality in health is a way to a universal comparability of degrees of inequality over time and across countries. This makes a difference to the problematic comparability of group-based (social class-based) measurement of inequality in health, which can be biased by differences in subjective labels of social classes and differences in their relative sizes (degrees of group's selectivity). In addition, there is a difficulty in attaching social-class labels to people who are not of working ages or do not work for other reasons.

Illsey and Le Grand (1987) computed Gini coefficient from distributions of deaths by age in real populations. Other researchers linked Gini coefficient and other measures of inter-individual inequality in age at death with the life table (Hanada, 1983, Silber, 1992, Wilmoth and Horiuchi, 1999, Anand and Nanthikesan, 2000, Anand et al., 2001).

Hicks proposed to use Gini coefficient to adjust average life expectancy for variability in order to construct the inequality-adjusted human development index (Hicks, 1997).

Gini coefficient has also been considered among other indices of inequality, as a measure of the rectangularization of survival curves in human populations (Wilmoth and Horiuchi, 1999). This approach is closely linked to the broader concept of studying historical evolution of human mortality, mortality compression, and limits of the human life span (Fries, 1980, Myers and Manton, 1984, Kannisto et al., 1994, Wilmoth and Lundström, 1996, Lynch and Brown, 2001).

The purpose of the present study is mostly a practical one. It is aimed at developing a toolkit for operating with Gini coefficient, similar to the one used for analyzing average life expectancy.

The first section of the study briefly presents a theoretical framework for measuring a degree of inter-individual inequality in length of life. It provides definitions and describes the basic properties of Gini coefficient, as well as four other measures of inequality in length of life. The second section considers a number of empirical trends in Gini coefficient and four other inequality measures, in order to see whether a judgement about the direction of change in inequality can be affected by a choice of indicator. The third section introduces a simple method for computation of Gini coefficient from discrete data of complete and abridged life tables. The fourth section presents new formulae for decomposing differences between two Gini coefficients, by age and cause of death, and a method for decomposing these differences by age, mortality and population group. Finally, the fifth section analyzes variations in Gini coefficient and in life expectancy across countries and over time.

So far, a similar research agenda has been completed only for one measure of variability in length of life, the interquartile range (Wilmoth and Horiuchi, 1999). We will show, however, that this measure has certain disadvantages, which could have undesirable consequences for empirical analyses.

## 2. Definitions and properties

### 2.1 The Lorenz curve (Note 1)

There is a vast amount of literature in economics on the measurement of inequality in income or wealth among households or individuals (Atkinson, 1970, Sen, 1973, Anand, 1983, Foster and Sen, 1997). The Lorenz (or concentration) curve is the most common device for a full description of distribution of income in a population. The Lorenz curve represents cumulative income share as a function of the cumulative population share.

Let  $f(x)$  be a population-density function of income  $x$ . Then the cumulative share of the population with income less or equal to  $x$   $F(x) = \int_0^x f(y)dy$  and the share of the

total income received by this part of the population is  $\Phi(x) = (1/\mu) \int_0^x yf(y)dy$ , where

the mean of the distribution  $\mu = \int_0^{\infty} yf(y)dy$ . The Lorenz curve as a function varies from

0 to 1 and is defined on the interval of variation of  $F(x)$  values  $[0,1]$ . In a situation of perfect equality for any income  $x$   $\Phi(x) = F(x)$ , the Lorenz curve is simply a diagonal, connecting points (0,0) and (1,1). If the horizontal axis corresponds to  $F(x)$  and the vertical axis corresponds to  $\Phi(x)$ , then the Lorenz curves for real income distributions would lie under the diagonal. The higher the variability in income across a population, the greater the divergence between the diagonal and the Lorenz curve.

Income distribution  $\mathbf{x}$  Lorenz-dominates income distribution  $\mathbf{y}$  if for any population share  $p$   $\Phi_x(p) \geq \Phi_y(p)$  (Anand, 1983). In this case the income distribution  $\mathbf{x}$  is considered as more equal (or less unequal) than the income distribution  $\mathbf{y}$ .

### 2.2 Application to the distribution of length of life

Applying this framework to mortality-by-age schedules, one can imagine a person's years lived from birth to death to be "income" and cumulative death numbers to be "population". Then the Lorenz curve can be constructed from the life table distribution by age at death. One can re-define the density and the distribution functions in terms of the standard life table functions as

$$f(x) = d(x)/l(0) \tag{1}$$

$$F(x) = 1 - l(x)/l(0) \tag{2}$$

$$\Phi(x) = \frac{1}{e(0)l(0)} \int_0^x td(t)dt \tag{3}$$

In practice, demographers have to operate with discrete data of life tables. The Lorenz curve can be defined on the basis of a complete life table as a set of points with

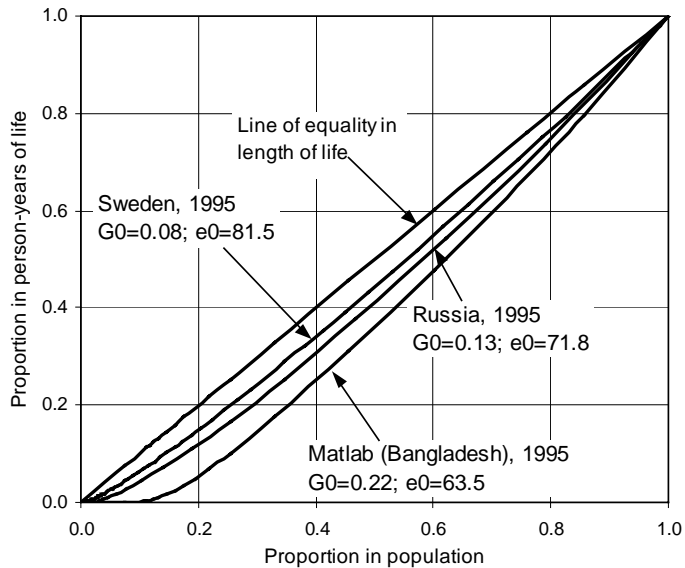
horizontal coordinates  $F_x = \frac{\sum_{t=0}^{x-1} d_t}{\sum_{t=0}^{\omega-1} d_t} = 1 - \frac{l_x}{l_0}$  and vertical coordinates

$$\Phi_x = \frac{\sum_{t=0}^{x-1} d_t \cdot \bar{t}}{\sum_{t=0}^{\omega-1} d_t \cdot \bar{t}} = \frac{T_0 - (T_x + xl_x)}{T_0}, \text{ where } \omega \text{ is the oldest age in the life table, } x$$

runs from 0 to  $\omega$ ,  $\bar{t}$  is the mean age at death of individuals dying between the exact ages  $t$  and  $t+1$ .

The situation of perfect equality takes place if all individuals die at the same age  $e_0$ . In this case, the line of perfect equality consists of only two end-points:  $F(x) = 0, \Phi(x) = 0$  for  $\forall x, x \neq e_0$  and  $F(x) = 1, \Phi(x) = 1$  for  $x = e_0$ .

Figure 1 shows an example of the Lorenz curves for three female life tables describing the very different female mortality patterns of Sweden, Russia, and Bangladesh (Matlab Report, 1996) in 1995. The Lorenz curve for Sweden dominates the curve for Russia, which in turn dominates the curve for Bangladesh.



**Figure 1:** Lorenz curves for three female populations with different average levels and age distributions of mortality.

Sources: Data for computations for Sweden are extracted from The Berkeley Mortality Database. Our own estimates are based on the original Goskomstat's data on deaths and population by age for Russia. Life table for Bangladesh is taken from Matlab Report (1996).

### 2.3 Gini coefficient

Various measures of inequality try to express in different ways a degree of inequality or variability as one number. Some of them are directly based on the Lorenz curve, and others are not. Gini coefficient is the best known and the most widely used measure of divergence based on the Lorenz curve. It is defined as an area between the diagonal and the Lorenz curve, divided by the whole area below the diagonal (equal to 1/2). Analytically it can be expressed as

$$G_0 = 1 - 2 \cdot \int_0^1 \Phi(p) dp, \text{ where } p = F(x) \quad (4)$$



Hanada (1983) showed that definition (4) together with (1)-(3) leads to

$$G_0 = 1 - \frac{1}{e(0)[l(0)]^2} \cdot \int_0^{\infty} [l(x)]^2 dx \quad (5)$$

Sometimes it is impossible to get mortality data for the full range of ages. For example, mortality data could be unreliable at infant ages or at old ages. Often, in studies combining inter-individual inequalities with inter-group (social class) inequalities in length of life (see section 4), group-specific data on mortality are available only for a limited range of ages (e.g. working ages). Therefore, one might want to measure the inequality in age at death for ages above 15 (denoted as  $G_{15}$ ) or between 20 and 65 (denoted as  $G_{20|65}$ ).

Formula (5) can be re-written for the range of ages  $[x, X]$

$$G_{x|X} = 1 - \frac{1}{e(x|X)[l(x)]^2} \cdot \int_x^X [l(t)]^2 dt, \quad (5a)$$

where the temporary life expectancy is  $e(x|X) = \frac{1}{l(x)} \int_x^X l(t) dt$  (Arriaga, 1984).

Gini coefficient varies between the limits of 0 (perfect equality) and 1 (perfect inequality). For a length-of-life distribution, it is equal to zero if all individuals die at the same age, and equal to 1 if all people die at age 0 and one individual dies at an infinitely old age.

There are several other ways to define Gini coefficient apart from the geometric definition (4). All of them are equivalent (Anand, 1983). The definition by Kendall and Stuart (1966) is especially helpful for understanding the nature of this measure.

$$G_0 = \frac{1}{2\mu} \int_0^{\infty} \int_0^{\infty} |x - y| f(x)f(y) dx dy \quad (6)$$

It suggests that Gini coefficient is simply a mean of absolute differences in individual ages at death (lengths of life) relative to the average length of life. If the population under consideration consists of  $l_0$  individuals, then the Gini coefficient is one-half of the average of absolute differences between all pairs of individual ages at death divided by the average length of life

$$G_0 = \frac{1}{2(l_0)^2} e_0 \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|. \quad (6a)$$

This expression can be re-written in terms of the standard life table functions as

$$G_0 = \frac{1}{2(l_0)^2} e_0 \sum_{x=0}^{\omega} \sum_{y=0}^{\omega} d_x \cdot d_y \cdot |\bar{x} - \bar{y}|, \quad (6b)$$

where  $\bar{x}$  and  $\bar{y}$  are the average ages at death for the elementary age intervals  $[x, x+1)$  and  $[y, y+1)$ , respectively.

Formula (7) is simple to understand, but is not easy to apply in practical calculations. In this respect, it is preferable to use Hanada's formula (5). As a construct, this formula is quite similar to the one for life expectancy  $e_0 = \frac{1}{l(0)} \cdot \int_0^{\infty} l(x) dx$ . The task is to estimate the area under the curve  $[l(x)]^2$  similarly to the area under curve  $l(x)$  for life expectancy. This similarity helps to find a simple way for calculation of  $G_0$  from discrete data (see section 3).

Formulae (6), (6a) and (7) make it clear that Gini coefficient is a mean-standardized measure. It varies from 0 to 1 and reflects relative inter-individual inequality. Such measures are also called indices. For some reasons one might be interested in absolute inter-individual differences in length of life. A respective measure, denoted  $G_0^{abs}$ , can be obtained from formulae (5) or (6) by removing life expectancy from the denominator (Note 2). It is equal to the average inter-individual difference in length of life and is measured in years. The Hanada's formula (5) yields

$$G_0^{abs} = G_0 \cdot e(0) = e(0) - \frac{1}{[l(0)]^2} \cdot \int_0^{\infty} [l(x)]^2 dx \quad (7)$$

The present paper focuses on the relative Gini coefficient, but sections 3 and 4 also show how to compute and decompose  $G_0^{abs}$ .

## 2.4 Basic properties of inequality measures

The following basic properties are desirable for any index of income inequality (Anand, 1983):

- (a) *population-size independence (or relativity principle)*, that is, the index does not change if the overall number of individuals changes with no change in proportions of individual incomes;
- (b) *mean or scale independence*, that is, the index does not change if everyone's income changes by the same proportion;
- (c) *Pigou-Dalton condition (or transfer principle)*, that is, any transfer from a richer to a poorer individual that does not reverse their relative ranks reduces the value of the index.

Satisfaction of these three conditions guarantees that an index of income inequality will correctly reflect the Lorenz-dominance (Anand, 1983). That is to say that in a comparison of two income distributions it will have a smaller value for the income distribution, which dominates another distribution.

Gini coefficient satisfies basic conditions (a) to (c) (Anand, 1983, Goodwin and Vaupel, 1985) and, therefore, correctly reflects the Lorenz-dominance among distributions of length of life. Note values of  $G_0$  in Figure 3 as an example.

Let us briefly consider a selection of other measures of inequality, which have already been applied elsewhere to distributions of length of life (Wilmoth and Horiuchi, 1999, Anand et al., 2001, Anand and Nathikesan, 2001), in light of basic properties (a) to (c).

The interquartile range (*IQR*) (see appendix 1 and Wilmoth and Horiuchi, 1999 for its exact definition) satisfies conditions (a) and (b), but does not satisfy the Pigou-Dalton condition (c). Indeed, an inter-individual transfer of years of life between two individuals will not change the *IQR* if both ages at death are either inside the quartile limits, both of them are higher than the upper limit, or both of them are lower than the lower limit. It means that the *IQR* ignores a change in distribution of deaths within certain age ranges if the total death numbers for each of these ranges do not change. For example, if  $x_{75}=60$  years (25,000 of 100,000 die at ages under 60) then it does not matter whether 20,000 die at ages under 15 and 5,000 die at ages from 15 to 59 or if 5,000 die at ages under 15 and 20,000 die at ages from 15 to 59.

The variance (*VAR*) and the standard deviation (*STD*) of the age at death (see Appendix 1 for definitions) satisfy conditions (a) and (c). However, these measures are

not standardized by the average life expectancy and, therefore, do not satisfy condition (b) of the mean- or scale- independence. Indeed, if all ages at death are multiplied by factor  $s$  then the variance of ages at death changes by factor  $s^2$  and *STD* changes by factor  $s$ . This means that the measure can change even if the distribution remains unchanged, but mean value changes (Note 3).

Unlike *VAR*, the variance of the logarithm of length of life (*VarLog*) (see appendix 1 for definition) is scale- and mean- independent, but it does not satisfy the Pigou-Dalton condition for ages above  $\hat{\mu}_e$  (where  $e$  is the base of the natural logarithm) (Anand, 1983).

The Theil entropy index ( $T$ ) is based on the notion of entropy in information theory (see appendix 1, Anand, 1983, and Theil, 1967). It satisfies all basic properties (a) to (c), but it is more difficult to understand or to interpret it in comparison to  $G$ , *IQR* or *STD*. Theil interprets  $T$  as "the expected information of a message which transforms population shares into income shares" (Theil, 1967, p. 95 cited by Anand, 1983, p. 309).

Different inequality measures are characterized by different sensitivity to changes in different sections of the length-of-life distribution. In practice, their sensitivity to changes in infant mortality is especially important. A. Atkinson constructed an inequality index with degrees of aversion to inequality expressed in an implicit form by a special parameter of aversion (Atkinson, 1970). Its minimum value of 0 means that all ages have the same weight in the inequality index. Higher values of the aversion parameter attach increasingly greater weights to earlier years of life (Anand et al, 2001).

Basic properties (a) to (c) and information on sensitivity to changes in tails of the length-of-life distribution suggest how a given measure of inequality will behave when applied to real mortality data. The following section shows that differences in properties of inequality measures have implications for empirical results.

### **3. Empirical trends in Gini coefficient and other measures of inequality and judgements about direction of changes in inequality**

Wilmoth and Horiuchi (1999) showed that for several industrialized countries, there is high correlation between long term trends in various measures of inequality. However, this result does not mean a full agreement between the directions of all temporal changes in various measures of inequality.

Section 1.5 indicates certain differences in properties of different measures of inequality. They suggest that it is theoretically possible that the same change in the mortality pattern can produce an increase in some measures of inequality and a decrease in other measures of inequality. Our earlier study showed an example of such a disparity for Russia in 1990-95 (Anand et al., 2001). It was demonstrated that if

mortality decreases at infant age and increases at adult ages, then Atkinson indices with high aversion parameters show an increase in inequality, while all other measures of inequality show a decrease. This section provides additional empirical evidence of the same nature by looking at similarities and dissimilarities among temporal changes in  $G_0$  and four other measures of inequality in length of life.

We consider trends in the selected measures of inequality for male populations of the USA and Russia since the 1950s. These countries and time periods have been selected as demonstrative ones after extensive exploratory analyses for nine industrialized countries over longer time periods.

Figure 2a suggests a remarkable similarity among proportional changes in the Gini coefficient, the standard deviation, the variance of log-life, the Theil entropy index, and the interquartile range for males in the USA from 1950 to 1995. All trends decline in response to transfers of deaths from younger to older ages and growing concentration of deaths at old ages.

Age-decompositions of changes (not shown here) demonstrate that differences between measures in the pace of temporal decline (steepest changes in *VarLog* and slowest changes in *IQR*) are mostly due to their varied sensitivity to decreasing infant mortality. *VarLog* continues to decrease in the 1960s in spite of mortality stagnation at adult ages and also in the 1980s-90s when all other measures of inequality stabilize. *T* also declines steeply in the 1950s due to declining infant mortality, but it becomes sensitive to changes at adult ages after 1960 when the number of infant deaths becomes low.

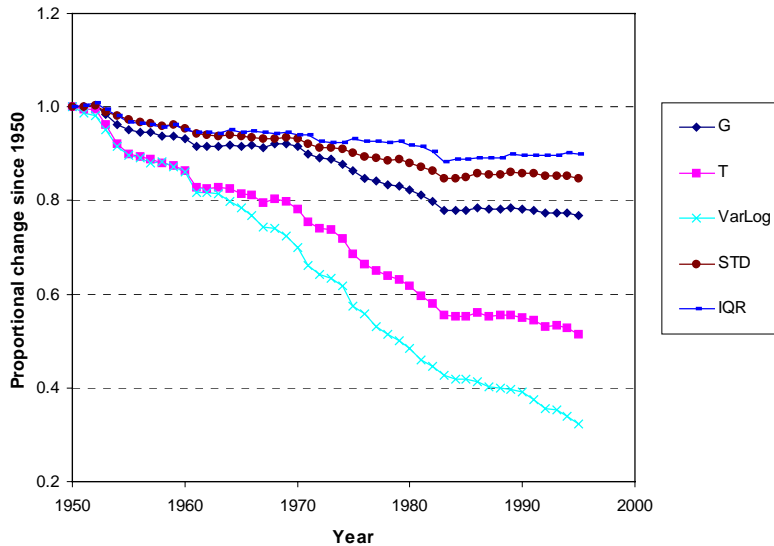
*IQR*,  $G_0$  and *STD* tend to stabilize in the 1960s due to mortality stagnation and also in the 1980s-90s. Reasons for the most recent stabilization are considered in section 5. *IQR* is also quite stable during the 1970s, while  $G_0$  and *STD* decrease. Once again, this difference is attributable to a varied sensitivity to mortality decline at young ages.

Russian mortality data provides an excellent opportunity for empirical testing of the inequality measures. This is due to the remarkable diversity among mortality trends for different age groups. In 1959-2000, mortality in infancy and childhood was decreasing, except for a short period of 1971-74. Mortality at ages from 15 to 69 was continuously increasing, except for a short period of sudden decline in 1985-87, and mortality at ages above 70 was mostly increasing at a slow pace (Shkolnikov, Meslé and Vallin, 1996). Life expectancy at birth increased from 1959 to 1964, and has continuously decreased since then, except 1985-87, when it increased.

Figure 2b confirms a high sensitivity of *VarLog* and *T* to declining infant mortality in comparison to *IQR*,  $G_0$  and *STD*. As in the USA, the trend in *VarLog* is mostly driven by infant mortality, being quite insensitive to huge changes at adult ages after 1985. *T* is also largely influenced by infant mortality, but since the mid-1960s it has reacted to

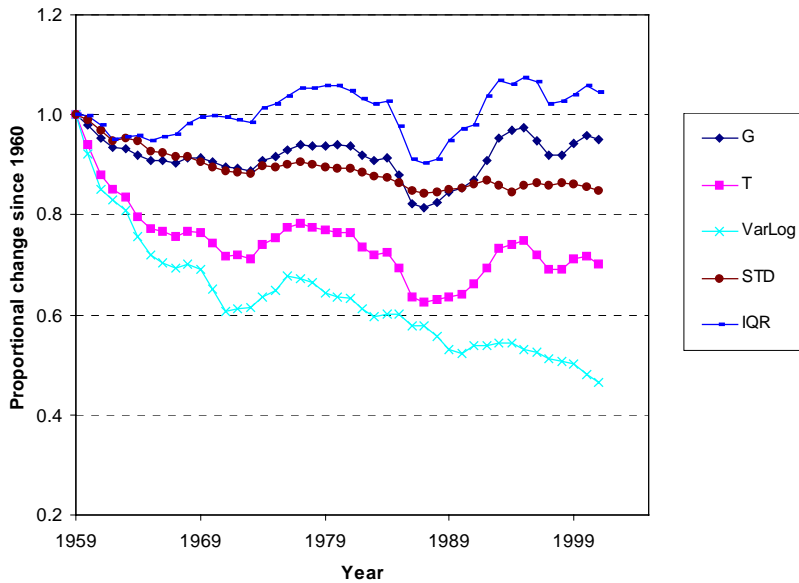
changes at adult ages. The trend in  $G_0$  is similar to that in  $T$ , with a less significant decrease from 1959 to 1965 and a greater magnitude of variations after 1985.

The trend in  $IQR$  deserves special attention. Interestingly,  $IQR$  increases between 1962 and 1970, while  $G_0$  and all other measures of inequality decrease. For a better understanding of this fact, it is useful to compare two time points with almost the same values of  $IQR$ ,  $x_{75}$  and  $x_{25}$ : 23.8, 54.5, and 78.3 years in 1959, respectively, and 23.7, 54.6, and 78.3 years in 1968, respectively. The number of life table deaths under age 55 is almost the same in 1959 as in 1968 (25,503 vs. 25,410 with radix 100,000). However, these deaths are distributed differently among ages under 55, with 26% of deaths under age 15 in 1959 vs. only 16% of deaths under age 15 in 1968. Age-distribution of deaths in 1965 is definitely more equal than that in 1959, but  $IQR$  remains invariant, unlike other measures. This is a real situation showing a consequence of the violation of the Pigou-Dalton condition (section 1.5) on empirical data.



**Figure 2a:** Proportional changes in Gini coefficient ( $G_0$ ), standard deviation (STD), variance of log-life (VarLog), Theil entropy index ( $T$ ), and interquartile range ( $IQR$ ) for males in the USA in 1950-97. (Level in 1950 is taken as 1).

Source: data for computation are extracted from The Berkeley Mortality Database (2001), originally constructed by the Office of the Actuary of the Social Security Administration.



**Figure 2b:** Proportional changes in Gini coefficient ( $G_0$ ), standard deviation (STD), variance of log-life (VarLog), Theil entropy index (T), and interquartile range (IQR) for males in Russia in 1959-2000. (Level in 1959 is taken as 1).

Source: data for computation are extracted from the original Goskomstat's annual files on deaths and mid-year population estimates.

The trend in *STD* shows a low variation across years relative to the starting year. In the case of Russia it appears that changes in the distribution of deaths by age lead to relatively minor changes in the sum of squares of deviations from the continuously declining average length of life. The coefficient of variation, equal to  $STD/e_0$ , experiences significantly greater magnitudes of proportional changes from 1959 to 2000 (not shown here).

Let us make another comparison of the length-of-life distributions for Russian men between 1965 and 1975. There is less inequality in 1975 compared to 1965 in terms of *VarLog* and *STD*, more inequality in 1975 compared to 1965 in terms of *IQR* and  $G_0$  and almost no difference between 1965 and 1975 in terms of *T*. This suggests that the choice of measure can change a conclusion about the direction of changes in inequality.

There is no reason to claim that one of all possible measures of inequality is the best. This claim is impossible in economics or statistics, and it is equally impossible in demography. Different measures emphasize different aspects of length-of-life

distributions. Relevance of their application depends on concrete purposes of analysis. On the other hand, some measures are used more widely than other ones because they fit better to typical situations and are easy for computation.

Theoretical considerations of the previous section, and empirical findings of this section, show that a violation of one of the basic properties (a) to (c) is a worrying signal. In this connection, one might prefer to use  $G_0$ ,  $T$  and  $VarLog$  rather than  $IQR$  or  $STD$ .

$VarLog$  is also not free from theoretical disadvantages (Anand, 1983), since it does not satisfy the Pigou-Dalton condition for the whole range of ages (see section 1.5). In addition, it is so sensitive to changes in infant mortality that it does not react significantly, even to large changes at adult ages if infant mortality continues to decline.

$T$  has no disadvantage in comparison to  $G_0$ , both from the theoretical or empirical sides. But it seems that this entropy-based measure is quite difficult for use in both understanding and interpretation.

Finally, in this study we favor Gini coefficient. It satisfies the basic properties in (a) to (c). Unlike many other inequality measures (shown and not shown here), it is not extremely sensitive to redistributions at early ages of life and reflects the changes at adult ages well.

Gini coefficient is an intuitively meaningful measure, which can be understood from its definitions (section 1.4) and from the analogy with income.

The following section shows how Gini coefficient can be computed from life table data.

#### 4. Computation of Gini coefficient from complete and abridged life tables

The availability of life tables for a population presumes an ability to estimate the

integral  $\int_0^{\infty} l(x)dx$  from discrete data. It would also help to estimate the integral

$$\int_0^{\infty} [l(x)]^2 dx \cdot$$

In a complete life table with  $l_0 = 1$  the life expectancy at birth is estimated as

$$e_0 = \sum_x \int_0^1 l(x+t)dt = \sum_x L_x = \sum_x [l_{x+1} + A_x(l_x - l_{x+1})].$$



For an elementary age interval  $[x, x+1)$ , parameter  $A_x$  is the average share of the interval lived by individuals, who die within the interval. These parameters are known from the life table  $A_x = \frac{(L_x / 1) - l_{x+1}}{l_x - l_{x+1}}$ .

Let us assume that the integral  $\int_0^1 [l(x)]^2 dx$  can be expressed in a similar way

$$\sum_x \int_0^1 [l(x+t)]^2 dt = \sum_x [(l_{x+1})^2 + \hat{A}_x ((l_x)^2 - (l_{x+1})^2)]. \quad (8)$$

Unknown parameters  $\hat{A}_x$  are to be estimated. For ages  $x > 0$  survival function  $l(x+t)$  can be reasonably well described by a parabola within the elementary age interval  $0 \leq t \leq 1$ . A parabola having the value  $l_x$  at  $t=0$  and the value  $l_{x+1}$  at  $t=1$  with the integral from 0 to 1 equal to  $L_x$  is

$$l(x+t) = l_x + (l_{x+1} - l_x)t + 6C_x (l_{x+1} - l_x)t(t-1), \quad (9)$$

where  $C_x = A_x - \frac{1}{2}$ .

It is possible then to determine a polynomial of the fourth degree for the function of our interest  $[l(x+t)]^2$  (see Appendix 2 for more details), and to derive the expression for  $\hat{A}_x$  by using (8)

$$\hat{A}_x = \frac{1 - \frac{2}{3}q_x + C_x(2 - q_x - \frac{6}{5}C_x)}{2 - q_x} \quad (10)$$

For a simple case with life table deaths evenly distributed within an elementary age interval  $A_x = \frac{1}{2}$  and  $C_x = 0$  and formula (10) yields

$$\hat{A}_x = \frac{1 - \frac{2}{3}q_x}{2 - q_x}.$$

It implies that  $\hat{A}_x = \frac{1}{2}$  if  $q_x = 0$  and  $\hat{A}_x = \frac{1}{3}$  if  $q_x = 1$ . If the probability of death is low (which is true for most of the ages in a complete life table) then the difference between  $\hat{A}_x$  and  $A_x$  is also very small. At old ages, where the probability of death is higher, the decrease in  $[l(x)]^2$  becomes considerably steeper than the decrease in  $l(x)$  and the deviation of  $\hat{A}_x$  from  $A_x$  becomes greater.  $\hat{A}_x$  tends to be smaller than  $A_x$ , consequently, a numerical integration (8) of the function  $[l(x)]^2$  by using the original life table  $A_x$  instead of "true" parameters  $\hat{A}_x$  would result in some underestimation of  $G_0$ .

Formula (10) is also valid for an abridged life table if for an elementary age interval  $[x, x+n)$  parameter  $A_x$  is defined as  $\frac{({}_nL_x/n) - l_{x+n}}{l_x - l_{x+n}}$  and, therefore varies between 0 and 1.

Formula (10) would not work in a proper way for  $x=0$  because during the first year of life  $l(x)$  falls more steeply than can be predicted by a quadratic polynomial. The use of the formula by J.Borgois-Pichat (1951) instead of a parabola (Appendix 3), solves the problem for age 0 and results in

$$\hat{A}_0 = A_0 \left( 1 - q_0 \frac{3 + 0.831 \cdot A_0}{2 + q_0} \right). \tag{11}$$

Let us, lastly, find a solution for an open-age interval. For many populations mortality data running up to the highest ages are not available. For example, in the WHO Mortality Database (2001), the last age group is 85+ for almost all countries and calendar years. Fortunately, The Berkeley Mortality Database (The Berkeley Mortality Database, 2001) provides 335 complete life tables for Japan, France, Sweden, and the USA with single-year age groups running up to 110. Mortality data by single-year age group for ages 85-110 provide the possibility to find a solution for the last age group 85+.

For an elementary age interval  $[x, x+n)$

$${}_nL_x = \int_x^{x+n} l(t)dt = n \cdot l_{x+n} + n \cdot A_x (l_x - l_{x+n}).$$

For the open-ended interval 85+  $l_\infty = 0$

and  $A_{85+}$  can be defined as  $A_{85+} = \frac{1}{l_{85}} \cdot \int_{85}^{\infty} l(x)dx = e_{85}$  (Note4).

In a similar way,  $\hat{A}_{85}$  can be defined as

$$\hat{A}_{85+} = \frac{1}{(l_{85})^2} \cdot \int_{85}^{\infty} [l(x)]^2 dx \cong \frac{1}{(l_{85})^2} \sum_{x=85}^{109} [(l_{x+1})^2 + \hat{A}_x ((l_x)^2 - (l_{x+1})^2)].$$

Having the values of  $l_x$  and  $A_x$  and computing  $\hat{A}_x$  from (10), one can obtain  $A_{85+}$ . The general relationship between  $\hat{A}_{85+}$  and  $A_{85+} = e_{85}$  can be estimated on the basis of the 335 life tables, by means of linear regression (Note 5). It returns the following equations:

$$\begin{aligned} \hat{A}_{85} &= -0.440 + 0.680 \cdot e_{85} \quad (\text{for women}), \\ \hat{A}_{85} &= -0.227 + 0.626 \cdot e_{85} \quad (\text{for men}) \end{aligned} \tag{12}$$

Formulae (10), (11), and (12) give a set of parameters  $\hat{A}_x$  for the numerical integration of the function  $[l(x)]^2$ . After this stage is completed it is easy to calculate  $G_0$  or  $G_0^{abs}$  according to formulae (5) and (7).

Table 1 shows the magnitudes of errors depending on the type of input data (complete life tables, abridged life tables with ages running up to 110 or abridged life tables with the last age 85+) and the method of computation (with  $A_x$  or with  $\hat{A}_x$ ) for Swedish life tables. These data for the years 1861, 1900, 1920, 1940, 1960, 1980, and 1995 are taken from The Berkeley Mortality Database (2001). Values of Gini coefficient computed from complete life tables with the last age 110, and adjusted  $\hat{A}_x$ , are considered as "exact" estimates. Abridged life tables and abridged life tables with the last age group 85+ are made from complete life tables in a conventional way. Table 1 suggests that if the data of complete life tables with the last age 110 are available, then it is not that important whether the original  $A_x$  or adjusted  $\hat{A}_x$  are used. Although in the former case where  $G_0$  estimates are systematically lower than the "exact" ones, the deviation is very small. As was mentioned before, the errors of approximating models are higher at infant and old ages, where functions  $l(x)$  and especially  $[l(x)]^2$  decline very steeply. Consequently, the errors of numerical integration are somewhat higher for historical populations with a high proportion of deaths in infancy, and for modern populations with a high proportion of deaths at advanced ages.

**Table 1:** Life expectancy at birth and different estimates of Gini coefficient for Sweden: computed from complete life tables, abridged life tables, abridged life tables with the last age group 85+ with and without modification of the life table  $A_x$ .

Year	Life expectancy $e_0$	Estimates of $G_0 \cdot 100$ from:						Errors in $G_0 \cdot 100$ estimates:				
		Complete LT with $\hat{A}_x$ ("exact" estimates)	Complete LT with $A_x$	Abridged LT with $A_x$	Abridged LT with $\hat{A}_x$	Abridged LT, last age 85+ with $A_x$	Abridged LT, last age 85+ with $\hat{A}_x$	Complete LT with $A_x$	Abridged LT with $A_x$	Abridged LT with $\hat{A}_x$	Abridged LT, last age 85+ with $A_x$	Abridged LT, last age 85+ with $\hat{A}_x$
		(1)	(2)	(3)	(4)	(5)	(6)	(1)-(2)	(1)-(3)	(1)-(4)	(1)-(5)	(1)-(6)
Males												
1861	45.03	38.140	38.087	37.995	38.152	37.890	38.152	0.053	0.145	-0.012	0.250	-0.012
1900	50.75	32.931	32.908	32.821	32.944	32.814	32.944	0.023	0.110	-0.013	0.117	-0.013
1920	57.42	26.658	26.647	26.557	26.668	26.533	26.668	0.011	0.101	-0.010	0.125	-0.010
1940	65.40	17.524	17.519	17.398	17.525	17.374	17.524	0.005	0.126	-0.001	0.150	0.000
1960	71.23	12.233	12.227	12.094	12.234	12.046	12.234	0.006	0.139	-0.001	0.187	-0.001
1980	72.78	11.128	11.122	10.998	11.133	10.923	11.132	0.006	0.130	-0.005	0.205	-0.004
1995	76.16	9.684	9.677	9.552	9.694	9.394	9.696	0.007	0.132	-0.010	0.290	-0.012
Females												
1861	48.78	35.436	35.403	35.319	35.457	35.309	35.457	0.033	0.117	-0.021	0.127	-0.021
1900	53.62	30.984	30.970	30.872	30.990	30.855	30.989	0.014	0.112	-0.006	0.129	-0.005
1920	60.11	24.627	24.620	24.516	24.627	24.475	24.626	0.007	0.111	0.000	0.152	0.001
1940	68.14	15.473	15.468	15.339	15.473	15.302	15.473	0.005	0.134	0.000	0.171	0.000
1960	74.87	10.382	10.376	10.237	10.387	10.129	10.385	0.006	0.145	-0.005	0.253	-0.003
1980	78.85	9.157	9.151	9.019	9.163	8.692	9.172	0.006	0.138	-0.006	0.465	-0.015
1995	81.45	8.337	8.331	8.192	8.339	7.630	8.381	0.006	0.145	-0.002	0.707	-0.044

Sources: Data for computations are extracted from The Berkeley Mortality Database (2001).

The values of  $G_0$  computed with  $A_x$  are relatively imprecise for abridged life tables, especially if the last age group is 85+ (Table 1). Markedly, the error has tended to increase quite significantly in the last five decades because the proportion of deaths occurring at ages above 85 has been increasing steeply.

Presented in Table 2 are the maximum relative deviations of five different estimates of  $G_0 \cdot 100$  from the "exact" estimates for 335 life tables taken from The Berkeley Mortality Database (for France, Japan, Sweden and the USA). It confirms the results of Table 1 on the basis of a large number of mortality schedules. Additionally, Table 2 shows that in the most recent years (low values of  $G_0$ ) the estimates computed from female abridged life tables, with the last age group 85+ using  $\hat{A}_x$ , are still shifted

slightly upwards from their "exact" values. This indicates that the use of  $\hat{A}_{85+}$  for the computation of  $G_0$  can not replace real mortality rates at ages above 85 if the proportion of life table deaths at age 85+ continues to increase. At present, the respective error is small, but it will increase with time and in the future it will be necessary to use mortality data for ages above 85.

In all cases the use of modified parameters  $\hat{A}_x$  reduce errors to a large extent and in all cases they become small. In order to re-check our prior results on the data, which had not been used to derive formulae (12) for estimating  $\hat{A}_{85+}$ , we made another comparison. First, we computed  $G_0 \cdot 100$  values from 89 complete life tables (from the Human Life-Table Database (2002)) with the last age group from 90 to 110 (mostly 100-105) for a diverse set of countries and years ("exact" estimates). Second, we computed two estimates of  $G_0 \cdot 100$  using  $A_x$  or  $\hat{A}_x$  from 89 abridged life tables with the last age 85+, corresponding to the complete life tables. For men, the average difference from the "exact" estimates of  $G_0 \cdot 100$  was 0.189 using  $A_x$ -estimates and only 0.014 using  $\hat{A}_x$ -estimates. For women, the equivalent figures were 0.291 and 0.026, respectively.

**Table 2:** *Maximum relative errors of the estimates of  $G_0 \cdot 100$  computed from complete life tables, abridged life tables, abridged life tables with the last age 85+ with and without modification of the life table by level of "exact"  $G_0 \cdot 100$ , in per cent.*

Ranges of "exact" values of $G_0 \cdot 100$	Number of life tables	Complete life tables, $A_x$	Abridged life tables, $A_x$	Abridged life tables, $\hat{A}_x$	Abridged life tables with the last age 85+, $\hat{A}_x$	Abridged life tables with the last age 85+, $\hat{A}_x$
<b>Males</b>						
< 15	145	-0.06	-1.40	0.06	-3.78	0.08
15 ≤ < 20	66	-0.05	-0.85	0.03	-1.08	0.03
20 ≤ < 25	38	-0.06	-0.59	0.04	-0.73	0.03
25 ≤ < 30	40	-0.08	-0.44	0.04	-0.49	0.04
30 ≤ < 35	29	-0.14	-0.40	0.06	-0.41	0.06
35 ≤	17	-0.27	-0.94	0.53	-0.94	0.53
<b>Females</b>						
< 10	72	-0.07	-1.78	0.09	-12.40	1.18
10 ≤ < 15	113	-0.04	-0.84	0.04	-1.30	0.03
15 ≤ < 20	35	-0.05	-0.60	0.04	-0.88	0.04
20 ≤ < 25	43	-0.07	-0.46	0.05	-0.57	0.04
25 ≤ < 30	41	-0.13	-0.40	0.05	-0.43	0.05
30 ≤	31	-0.14	-0.38	0.05	-0.39	0.05

Sources: Data for computations are extracted from the Human Life-Table Database (2002).

## 5. Decomposition of a difference between two values of Gini coefficient

### 5.1. Age

When analyzing changes in life expectancy over time or its variations across countries it is useful to be able to decompose observed differences by age and cause of death. This gives one the opportunity of linking variations in overall life expectancy with variations in elementary age-specific mortality rates. For a similar reason the idea of decomposition of differences between two Gini coefficients arises. The age-components would show to what extent differences in elementary mortality rates at different ages influence the overall difference in degrees of inequality in length of life.

The discrete method for the decomposition of a difference between two life expectancies by age was independently developed in the 1980s by three researchers from Russia, the USA, and France (Andreev, 1982, Arriaga, 1984, Pressat, 1985). The formula of decomposition by E.Andreev is exactly equivalent to that by R.Pressat. The formula by E.Arriaga is written in a slightly different form, but is actually equivalent to the formula by Andreev and Pressat (Shkolnikov et al., 2001). A continuous version of the method of decomposition by age was developed by Pollard (1982).

All of these methods are based on the idea of standardization or replacement (Kitagawa, 1964). If there are two populations under consideration then mortality rates of the first population are to be replaced in an age-by-age mode by mortality rates of the second population, or vice versa (Andreev, Shkolnikov, Begun, 2002). The contribution of a particular age group  $x$  to the overall difference in life expectancy can be computed as the difference between life expectancy of the first population and the life expectancy of the first population after replacement of the mortality rate at age  $x$  by the respective mortality rate of the second population.

First, we apply this general algorithm to a difference between  $e_0$  values and demonstrate that it leads to the conventional formula of decomposition. We then apply the same approach again to develop the formulae for decomposition of the difference between  $G_0$  values.

Let  $\mu^{[x]}$  be the force of mortality function equal to the force of mortality of the second population  $\mu'(t)$  if  $t \leq x$  and equal to the force of mortality of the first population  $\mu(t)$  if  $t > x$ . Then the difference in life expectancy at birth produced by replacement of force of mortality from 0 to  $x$  is

$$\delta_{0,x} = e_0(\mu^{[x]}) - e_0 = (L'_{0|x} - L_{0|x}) + (l'_x - l_x) \cdot e_x, \quad (13)$$

where  $L_{0|x} = \int_0^x l(t)dt$ . The first additive term in the expression is the effect of replacement at ages under  $x$ ; the second additive term is the effect of replacement at ages under  $x$  on length of life after age  $x$ . If the range of ages is divided into  $n$  intervals  $[x_i, x_{i+1})$  then the overall difference between the two life expectancies can be decomposed into age-specific contributions as

$$e'_0 - e_0 = \sum_{i=0}^{n-1} (\delta_{0,x_{i+1}} - \delta_{0,x_i}) = \sum_{i=0}^n \delta_i \quad (14)$$

$\delta_i$  can be regarded as a contribution of an elementary age interval  $[x_i, x_{i+1})$  to the overall difference between life expectancies at birth. Using (14) and (13) we easily come to the conventional formula of decomposition by Andreev and Pressat

$$e'_0 - e_0 = \sum_{i=0}^n [l'_{x_i} (e'_{x_i} - e_{x_i}) - l'_{x_{i+1}} (e'_{x_{i+1}} - e_{x_{i+1}})].$$

Components  $\delta_i$  can be presented in a more general way as

$$\delta_i = e_0 (M^{[x_{i+1}]} - M^{[x_i]}), \quad (15)$$

where  $M^{[x_i]}$  is a vector of age-specific mortality rates with elements  $m'_x$  for  $x \leq x_i$  and  $m_x$  for  $x > x_i$ . In fact, formula (15) can be considered as a general procedure for decomposition by age of a difference between two aggregate measures (Andreev, Shkolnikov and Begun, 2002). It determines a stepwise replacement of one mortality schedule by another one, beginning from the youngest and proceeding to the oldest age group. Discussion about this and other orders of replacement in respect to the age-decompositions can be found elsewhere (Pollard, 1988, Das Gupta, 1994 and 1999, Horiuchi, Wilmoth and Pletcher, 2001, Andreev, Shkolnikov and Begun, 2002).

A formula for age-components for the difference between two  $G_0$  values can be developed by using definition (5) in a way similar to (13)-(14). The difference induced by mortality replacement at age  $x$  and younger ages can be expressed as

$$\varepsilon_{0,x} = G_0(M^{[x]}) - G_0 = \frac{\theta_0}{e_0} - \frac{\theta'_{0|x} + \theta_x (l'_x)^2}{e'_{0|x} + e_x l'_x} \quad (16)$$

where  $\theta_x = \frac{1}{(l_x)^2} \int_x^\infty [l(t)]^2 dt$ ,  $\theta_{0,x} = \int_0^x [l(t)]^2 dt$  and  $e_{0,x} = \int_0^x l(t) dt$ .

The decomposition of the difference in between two Gini coefficients by age group similar to (14) is

$$G'_0 - G_0 = \sum_{i=0}^{n-1} (\varepsilon_{0,x_{i+1}} - \varepsilon_{0,x_i}) = \sum_{i=0}^n \varepsilon_i \quad (17)$$

and a general procedure for the computation of age-specific components of the difference is

$$\varepsilon_i = G_0(M^{[x_{i+1}]} - M^{[x_i]}) \quad (18)$$

Similar formulae for the age-components of the difference between two  $G_0^{abs}$  values are given in appendix 4.

Formulae (16) and (17) allow a difference between two Gini coefficients to be split according to age groups. Similar to life expectancy (Andreev, 1982, Pressat, 1985), the results of decomposition are not exactly the same for the difference  $G'_0 - G_0$  in comparison to the difference  $G_0 - G'_0$ . That is to say that it does matter which mortality schedule is the basic one, which has to be replaced by another one. A conventional way to avoid this problem is to perform the decomposition (17) twice, and then to average the resulting age-specific components. In the present paper this technique is used in all decompositions.

Formulae (16) and (17) are analytical expressions, permitting a direct computation. Numerical integration for values of  $\theta_x$  and  $\theta_{0,x}$  can be completed by using the technique developed in the previous section (usage of the adjusted  $\hat{A}_x$  instead of the life table  $A_x$ ).

Procedure (18) can also be used directly for computation instead of formulae (16) and (17). Section 4.3 demonstrates that this procedure can be used for other types of decomposition (for example, by age and social group), where analytical expressions for components are not easily available.



**Table 3:** Age-specific contributions to the increase in life expectancy at birth and the decrease in Gini coefficient from 1900 to 1995: USA, men\*

Age group	Components of difference in $e_0$		Components of difference in $G_0 \cdot 100$	
	Absolute	%	Absolute	%
All ages	25.96	100.0	-24.02	100.0
0	8.40	32.3	-10.99	45.7
1-4	4.25	16.4	-5.50	22.9
5-14	1.76	6.8	-2.15	9.0
15-24	1.87	7.2	-2.00	8.3
25-39	2.94	11.3	-2.61	10.9
40-64	4.30	16.6	-1.91	7.9
65+	2.44	9.4	1.14	-4.7

\*  $e_0(1900)=46.4$ ,  $e_0(1995)=72.73$   $G_0(1900)=36.73$ ,  $G_0(1995)=12.71$

Sources: Data for computations are extracted from The Berkeley Mortality Database (2001). Life tables for the United States are based on those constructed by the Office of the Actuary of the Social Security Administration.

Table 3 shows the results of the decomposition of increase in life expectancy at birth and of the decrease in the Gini coefficient in the USA between 1900 and 1995. The total increase in life expectancy at birth is 26 years for men and 30 years for women, and the total decrease in  $G_0 \cdot 100$  is about 24 for both sexes. 55% of the overall increase in life expectancy is due to a decrease in mortality at ages 0-14. A 35% increase in life expectancy for men and 39% for women is due to a decrease in mortality at ages 15-64, and a further 9% increase for men and 17% for women is due to a decrease in mortality at ages 65 and older. The overall decrease in Gini coefficient is distributed somewhat differently. The proportion of the decrease due to the youngest age group 0-14 is higher than that for life expectancy (78% for men and 70% for women). The proportion of the medium age group is somewhat lower (27% for men and 37% for women), and the oldest age group produces a negative contribution of – 5%.

Gini coefficient, as a measure of dispersion, is more heavily influenced by the tails of the distribution than the life expectancy as a mean value with the null aversion to inequality (see section 1.5). Gini coefficient decreases when life table deaths concentrate around the average age at death. Historical reduction of infant mortality caused great equalization of ages at death. Indeed, in the course of mortality evolution, infants were achieving more years of life by avoiding deaths than adults or old people

by avoiding deaths. This is similar to an income redistribution, with the poorest gaining more additional income than the rich.

### 5.2. Age and cause of death

With the help of definition (5), formula (16) can be re-written without  $e_{0|x}$  and  $\theta_{0|x}$  as

$$\varepsilon_{0,x} = \frac{\theta_0}{e_0} - \frac{\theta'_0 - \theta'_x(l'_x)^2 + \theta_x(l'_x)^2}{e_0 - l'_x e'_x + e_x l'_x} \quad (19)$$

The following relations are true for a small  $\Delta x$ :

$$l_{x+\Delta x} = l_x(1 - \mu_x \Delta x), \quad e_{x+\Delta x} = e_x - (1 - \mu_x e_x) \Delta x, \quad \theta_{x+\Delta x} = \theta_x - (1 - 2\mu_x \theta_x) \Delta x.$$

Applying (17) and (19) to a small age interval  $[x, x + \Delta x]$  after some transformations, one can yield

$$\begin{aligned} \lambda_x &= \frac{\varepsilon_{x,x+\Delta x}}{\Delta x} = \frac{(\mu_x - \mu'_x)l'_x}{[e'_0 + l'_x(e_x - e'_x)]^2} \{e_x[\theta'_0 + (l'_x)^2(\theta_x - \theta'_x)] - 2l'_x\theta_x[e'_0 + l'_x(e_x - e'_x)]\} = \\ &= (\mu_x - \mu'_x)\eta_x. \end{aligned} \quad (20)$$

Integrating (20) from  $x_i$  to  $x_{i+1}$  yields

$$\varepsilon_{x_i, x_{i+1}} = \int_{x_i}^{x_{i+1}} \lambda_t dt = \sum_{j=1}^m \int_{x_i}^{x_{i+1}} (\mu_t - \mu'_t)\eta_t dt \cong (m_{x_i|x_{i+1}} - m'_{x_i|x_{i+1}}) \cdot \int_{x_i}^{x_{i+1}} \eta_t dt.$$

If there are  $m$  causes of death, then  $\mu_x = \sum_{j=1}^m \mu_{x,j}$  and

$$\varepsilon_{x_i, x_{i+1}} \cong \left( \int_{x_i}^{x_{i+1}} \eta_t dt \right) \cdot \sum_{j=1}^m (m_{x_i|x_{i+1},j} - m'_{x_i|x_{i+1},j}).$$

If in each age group  $[x_i, x_{i+1})$   $m_{x_i|x_{i+1}} \neq m'_{x_i|x_{i+1}}$  then cause-specific components within

this age group are simply

$$\varepsilon_{x_i, x_{i+1}, j} = \frac{m_{x_i|x_{i+1}, j} - m'_{x_i|x_{i+1}, j}}{m_{x_i|x_{i+1}} - m'_{x_i|x_{i+1}}} \cdot \varepsilon_{x_i, x_{i+1}},$$

otherwise a numerical integration would be necessary to compute cause-specific components according to

$$\varepsilon_{x_i, x_{i+1}, j} \cong \left( \int_{x_i}^{x_{i+1}} \eta_t dt \right) \cdot (m_{x_i|x_{i+1}, j} - m'_{x_i|x_{i+1}, j}).$$

Similar formulae for age- and cause-specific components of a difference between two  $G_0^{abs}$  values are given in appendix 4.

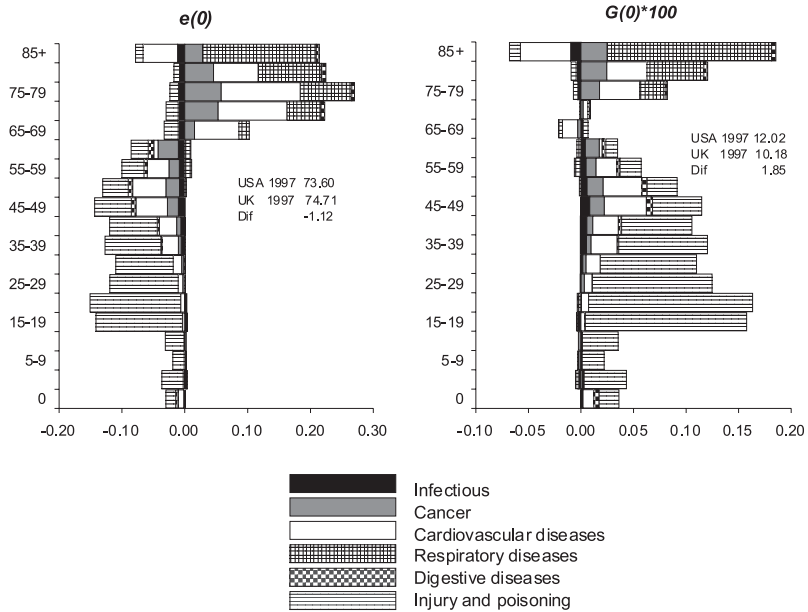
A comparison between the USA and the UK male life expectancies at birth and the USA and the UK Gini coefficients in the year 1997 is given as an example of decomposition by age and cause of death (Figure 3). The life expectancies of men are very similar in both countries. The difference is only one year in favor of the UK (or 1.4%). However, there is a significant 16% difference in the Gini coefficients with less inequality in length of life in the UK.

Figure 3 shows age- and cause-specific components of these differences. The advantage of the UK in male life expectancy (left panel of Figure 3) is mostly due to lower mortality rates by external causes of death (accidents and violence) at ages from 15 to 50 and, to some extent, to lower mortality rates by circulatory disease and cancers at ages from 40 to 59. However, this advantage is almost balanced by the effects of lower mortality in the USA at ages above 65 by circulatory and respiratory diseases and cancers.

The weight of external causes of death at young adult ages is higher for the UK-USA difference in the Gini coefficients (right panel of Figure 3) than that for the difference in life expectancies at birth. In addition, low mortality at old ages increases the level of the Gini coefficient in the USA in comparison to the UK.

Elimination of causes of death is another method for analyzing the influence of causes of death on life table measures. A conventional procedure of building the "associated" single decrement life table can be applied (Chiang, 1968, Preston et al., 2001). Gini coefficient can be computed from this table with modified  $\hat{A}_x$  (as described in section 1).

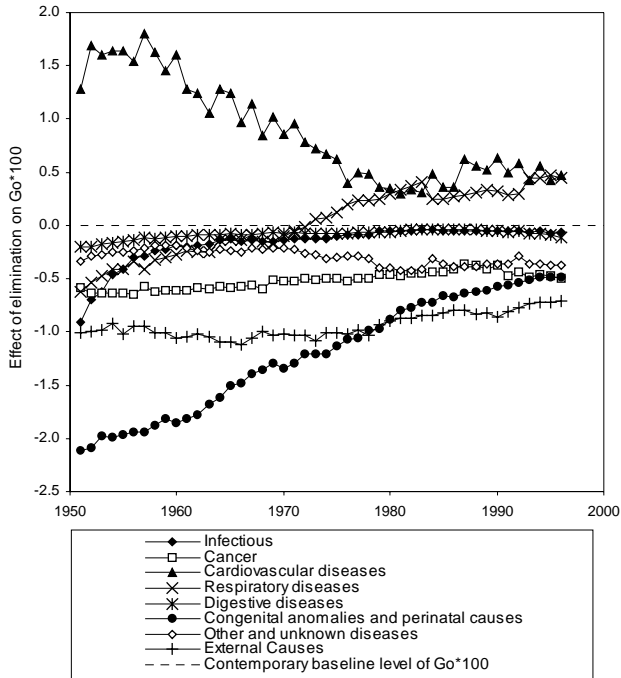
The elimination of causes of death produces a decrease or an increase in Gini coefficient, depending on ages affected by respective causes. Figure 4 shows temporal changes in the effects of elimination of leading classes of causes of death on  $G_0 \cdot 100$



**Figure 3:** *Decompositions of the differences in life expectancy at birth and in Gini coefficient between the UK and the USA by age and cause of death: male populations, 1997.*

Sources: Data for computations are extracted from the WHO Mortality Database (2001)

for women in the UK in 1951-1996. For the majority of causes of death, elimination leads to a decrease in  $G_0$ . There are two exceptions: cardiovascular diseases during the whole period, and respiratory diseases after 1973. Their elimination increases  $G_0$  because the distribution of life table deaths by age becomes more unequal. A reason for the change in the role of respiratory diseases in the 1970s is clear. This class of causes of death has transformed from an important cause of infant and child death into a cause of death of old people.



**Figure 4:** *Effects of elimination of causes of death on  $G_0$  for women in the UK in 1951-1996*

Sources: Data for computations are extracted from the WHO Mortality Database (2001)

In the 1950s, the elimination of perinatal causes and congenital anomalies could have led to the greatest decrease in  $G_0$  among other causes of death. By the 1990s, the effect of this major cause of infant death had been greatly reduced. The same has happened to other common causes of mortality in childhood (infectious and respiratory diseases). The elimination effects of cancers and external causes of death have been relatively stable in time.

At present, eliminating external causes of death, can produce the greatest decrease in  $G_0$  for women in the U.K.

### 5.3. Age, mortality and population structure

Inter-group differentials in mortality influence  $G_0$  if they affect the age pattern of mortality of the whole population. If a change of the inter-group mortality gaps and of the population-weights of the groups makes the age-distribution of deaths in the whole population less uneven, then  $G_0$  decreases. The decomposition of Gini coefficient by population group is an opportunity to link inter-individual and inter-group variations in length of life.

Additional dimensions in the data suggest that there are many different ways to replace group- and age-specific mortality rates and composition by group of one population by respective rates and composition by group of the other population. For example, one can make a replacement of mortality rates by age within each population group, or replace group-specific mortality rates within one age group. Generally speaking, all replacement schemes are equally acceptable and, therefore, a general algorithm for decomposition of the difference in aggregate measures should be based on the averaging of effects produced by all possible combinations of replacements (Das Gupta, 1994, 1999).

However, a concrete formulation of the decomposition "task" leads to a concrete replacement scheme. For example, it might be of interest to estimate impacts of mortality and population structure by group at each age. This implies the problem of splitting each age-component of the overall difference between  $G_0$  values into additive components related to mortality rates and population composition in respective age groups. This can be done after some modification of the algorithm of linear replacement determined by (15) and (18).

Let  $M = \|m_{ij}\|$  be a matrix of mortality rates by age group  $i$  and population group  $j$  and  $P = \|p_{ij}\|$  be a matrix of the weights of groups in the overall population of age group  $i$  ( $\sum_j p_{ij} = 1$  for every age group  $i$ ). For a given age group  $k$  the age-specific mortality rate for two populations under consideration are  $m_k = \sum_j p_{kj} \cdot m_{kj}$  and  $m'_k = \sum_j p'_{kj} \cdot m'_{kj}$ .

Let us define a "partly replaced" matrix of mortality rates  $M^{[k]}$  consisting of elements  $m'_{ij}$  for  $i \leq k$  and elements  $m_{ij}$  for  $i > k$ . A corresponding matrix of population weights with replaced rows (age groups) up to the age group  $k$   $P^{[k]}$  is defined in a similar way.

According to the general logic of replacement (18), the component of inter-population difference in  $G_0$  produced by age group  $k$  is

$$\varepsilon_k = G_0(\mathbf{M}^{[k]}, \mathbf{P}^{[k]}) - G_0(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k-1]}).$$

We consider two possible paths for a transition  $(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k-1]}) \rightarrow (\mathbf{M}^{[k]}, \mathbf{P}^{[k]})$ :

$$\begin{aligned} &(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k-1]}) \rightarrow (\mathbf{M}^{[k]}, \mathbf{P}^{[k-1]}) \rightarrow (\mathbf{M}^{[k]}, \mathbf{P}^{[k]}) \text{ or} \\ &(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k-1]}) \rightarrow (\mathbf{M}^{[k-1]}, \mathbf{P}^{[k]}) \rightarrow (\mathbf{M}^{[k]}, \mathbf{P}^{[k]}). \end{aligned}$$

Accordingly, it is possible to get two versions of the components due to mortality rates (M-effects  $\varepsilon_k^{1,M}, \varepsilon_k^{2,M}$ ) and to population composition (P-effects  $\varepsilon_k^{1,P}, \varepsilon_k^{2,P}$ ) for the age group  $k$ :

$$\varepsilon_k = \varepsilon_k^{1,M} + \varepsilon_k^{1,P} = [G_0(\mathbf{M}^{[k]}, \mathbf{P}^{[k]}) - G_0(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k]})] + [G_0(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k]}) - G_0(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k-1]})]$$

or

$$\varepsilon_k = \varepsilon_k^{2,P} + \varepsilon_k^{2,M} = [G_0(\mathbf{M}^{[k]}, \mathbf{P}^{[k]}) - G_0(\mathbf{M}^{[k]}, \mathbf{P}^{[k-1]})] + [G_0(\mathbf{M}^{[k]}, \mathbf{P}^{[k-1]}) - G_0(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k-1]})].$$

The final M-effects and P-effects for age group  $k$  can be obtained by averaging

$$\varepsilon_k^M = \varepsilon_k^{1,M} + \varepsilon_k^{2,M} = \frac{1}{2} \{ [G_0(\mathbf{M}^{[k]}, \mathbf{P}^{[k]}) - G_0(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k]})] + [G_0(\mathbf{M}^{[k]}, \mathbf{P}^{[k-1]}) - G_0(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k-1]})] \} \quad (21)$$

$$\varepsilon_k^P = \varepsilon_k^{1,P} + \varepsilon_k^{2,P} = \frac{1}{2} \{ [G_0(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k]}) - G_0(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k-1]})] + [G_0(\mathbf{M}^{[k]}, \mathbf{P}^{[k]}) - G_0(\mathbf{M}^{[k]}, \mathbf{P}^{[k-1]})] \} \quad (22)$$

Expressions (21) and (22) allow for distinguishing between contributions of mortality and population composition within every age-contribution  $\varepsilon_k$ .

It might be of additional interest to split M-effects according to particular population groups. To do so, we should re-define the replacement procedure for M-transitions  $(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k-1]}) \rightarrow (\mathbf{M}^{[k]}, \mathbf{P}^{[k-1]})$  and  $(\mathbf{M}^{[k-1]}, \mathbf{P}^{[k]}) \rightarrow (\mathbf{M}^{[k]}, \mathbf{P}^{[k]})$ . In our prior consideration it was very simple: row  $k$  was to be replaced entirely. However, to obtain the effect of the mortality rate in the particular population group  $j$  for age group  $k$ , two additional steps should be completed:

1) Computation of all effects of the replacement of the element  $m_{kj}$  by  $m'_{kj}$  in various combinations with  $m_{kl}$  or  $m'_{kl}$  for population groups  $l \neq j$ . If the number of population groups is  $L$  then the number of different replacements would be  $2^{L-1}$ .

2) Computation of each  $(k,j)$ -effect by the averaging of all  $l$ -effects for each  $j$ .

For example, if we make M-replacement for age 20 and there are two population groups 1 and 2, then the effect of the mortality rate in population group 1 at age 20 would be

$$\begin{aligned} \varepsilon_{20}^{M(P1)} &= \frac{1}{2} \left\{ \frac{1}{2} [G_0(\|m'_{20,1}, m'_{20,2}\|, P^{[20]}) + G_0(\|m'_{20,1}, m'_{20,2}\|, P^{[20]})] - G_0(M^{[19]}, P^{[20]}) \right\} + \\ &+ \frac{1}{2} \left\{ \frac{1}{2} [G_0(\|m'_{20,1}, m'_{20,2}\|, P^{[19]}) + G_0(\|m'_{20,1}, m'_{20,2}\|, P^{[19]})] - G_0(M^{[19]}, P^{[19]}) \right\} \end{aligned}$$

The equivalent M-effect for age 20 and population group 2 is

$$\begin{aligned} \varepsilon_{20}^{M(P2)} &= \frac{1}{2} \left\{ [G_0(\|m'_{20,1}, m'_{20,2}\|, P^{[20]}) + G_0(\|m'_{20,1}, m'_{20,2}\|, P^{[20]})] - G_0(M^{[19]}, P^{[20]}) \right\} + \\ &+ \frac{1}{2} \left\{ [G_0(\|m'_{20,1}, m'_{20,2}\|, P^{[19]}) + G_0(\|m'_{20,1}, m'_{20,2}\|, P^{[19]})] - G_0(M^{[19]}, P^{[19]}) \right\} \end{aligned}$$

Finally, one should keep in mind that the results of decomposition depend on permutations of populations. So, decomposition should be run twice.

Table 4 shows the educational composition of Russian men by age according to the censuses of 1979 and 1989. In 1979 the proportion of people with the lowest educational attainment was much higher at ages over 40 than at younger ages, by 1989 the borderline had moved up to age 50. In general, between 1979 and 1989, the educational composition improved significantly in terms of the proportions of university and secondary levels of education compared to the proportion of low educational level. Three processes contributed to this favorable change: production of highly educated people by the educational system, the natural replacement of older generations with lower educational levels by younger people with higher educational levels, and, to some extent, migration into Russia of people with relatively high levels of education from other parts of the USSR.

Table 5 suggests that between 1979 and 1989, life expectancy (20-64) increased and that inequality in age at death, measured by Gini coefficient, decreased. As we know, these changes occurred in the second half of the 1980s and can probably be attributed to Gorbachev's anti-alcohol campaign of 1985 (Shkolnikov et al., 1996). Improvements were the greatest for men with secondary education, followed by those with university education. In the group with a low education, achievements in life expectancy (20-64) and a decrease in Gini coefficient (20-64) were very modest. Interestingly, for the whole population, the increase in life expectancy (1.4 years) and the decrease in Gini coefficient (-2.4) were substantially greater than those for each of



the educational groups. This seeming paradox is due to the additional positive effect of change in educational composition shown in our earlier study (Shkolnikov et al., 1998).

**Table 4:** *Educational composition of the Russian male population by age in 1979 and 1989.*

Age group	University education	Secondary education	Lower education
	1979		
20-24	0.089	0.584	0.327
25-29	0.136	0.489	0.376
30-34	0.175	0.438	0.387
35-39	0.156	0.321	0.522
40-44	0.145	0.258	0.596
45-49	0.091	0.161	0.748
50-54	0.084	0.172	0.743
55-59	0.113*	0.215*	0.672*
60-64	0.085	0.173	0.743
1989			
20-24	0.109	0.780	0.111
25-29	0.160	0.734	0.106
30-34	0.165	0.679	0.155
35-39	0.176	0.591	0.233
40-44	0.206	0.504	0.290
45-49	0.173	0.374	0.453
50-54	0.159	0.299	0.542
55-59	0.100	0.194	0.706
60-64	0.094	0.194	0.712

Sources: Computed from the original statistical tables on population by age and educational status (Interstate Statistical Committee of CIS, 1996).

\* Proportions of people with high levels of education decline with age. The male age group 55-59 in 1979 is an exception. This group is very small in comparison with neighboring age groups due to heavy losses in the Second World War. Hence, it faced relatively low competition rate for entering higher school levels after the war.

**Table 5:** *Life expectancy and Gini coefficient for the range of ages from 20 to 64 in the Russian male population in 1979 and 1989\*.*

Population group	$e_{20 65}$			$G_{20 65} \cdot 100$		
	1979	1989	Difference	1979	1989	Difference
Total population	37.95	39.30	1.35	13.66	11.23	-2.43
University education	41.19	42.09	0.90	7.70	5.99	-1.72
Secondary education	38.45	39.51	1.06	12.69	10.87	-1.82
Lower education	36.70	37.04	0.34	16.04	15.57	-0.47

\* Detailed information about the data sources and the quality of the estimates of mortality by educational level can be found in Shkolnikov et al. (1998).

Sources: Computed from the original Goskomstat's tables on deaths and population by age and educational status.

A widening of inter-group differences in life expectancy of Russian men in 1979-89 coincides with a substantial decline in  $G_{20|65}$  (Table 5). This illustrates a difference in the meanings of inter-individual inequality in length of life and length-of-life differentials across social groups.

The advantage of the group of Russian men with university education is much more pronounced in Gini coefficient than in life expectancy. Indeed, a gap between university and low (lower than secondary) education in the Gini coefficient (20-64) constitutes 61% of its value in 1979 and 85% of its value in 1985. The equivalent percentages for life expectancy (20-64) are 11% in 1979 and 13% in 1989. Hence, the educational gradient is more pronounced in the variability than in the average age at death.

Table 6 highlights the "anatomy" of the increase in life expectancy (20-64) and decrease in Gini coefficient (20-64) for Russian men in 1979-89. Age-specific components are divided into effects of mortality (M-effects) and effects of educational composition (P-effects). The role of compositional effect in the improvements of the 1980s is very significant since its magnitude is almost the same as that of the mortality effect, especially for Gini coefficient.

Overall, there is a remarkable similarity between the structures of changes in life expectancy (20-64) and Gini coefficient (20-64). For the latter, the weight of components related to ages under 40 is somewhat higher than that for the former. For both measures the highest effects are related to ages from 30 to 45. The most significant contributions to overall improvement in both measures are produced by changes in mortality in the group with secondary education. Although mortality decline in the group with low education is very small (Table 5), its contribution to overall

improvement is greater than that for university education because low education has a higher weight in the population (Table 6).

**Table 6:** *Components of changes between 1979 and 1989 in life expectancy and Gini coefficient for the range of ages from 20 to 64 in the Russian male population.*

Age	Components produced by changes in mortality rates (M-effects)				Total	Components due to changing educational composition (P-effects)	Total
	University education	Secondary education	Low education	Total			
<i>e</i> <sub>20 65</sub> in years							
20-24	0.009	0.040	-0.006	0.043	0.078	0.120	
25-29	0.010	0.047	-0.016	0.042	0.125	0.167	
30-34	0.014	0.085	0.032	0.131	0.099	0.230	
35-39	0.017	0.068	0.014	0.099	0.104	0.203	
40-44	0.019	0.071	0.058	0.149	0.109	0.258	
45-49	0.012	0.040	0.013	0.064	0.072	0.136	
50-54	0.014	0.042	0.050	0.107	0.044	0.151	
55-59	0.010	0.015	0.055	0.080	-0.004	0.076	
60-64	0.003	0.003	0.001	0.008	0.001	0.008	
Total	0.109	0.412	0.201	0.722	0.628	1.349	
<i>G</i> <sub>20 65</sub>							
20-24	-0.020	-0.088	0.012	-0.095	-0.173	-0.268	
25-29	-0.022	-0.100	0.033	-0.089	-0.266	-0.354	
30-34	-0.028	-0.170	-0.065	-0.263	-0.198	-0.461	
35-39	-0.031	-0.128	-0.025	-0.185	-0.195	-0.380	
40-44	-0.033	-0.123	-0.100	-0.257	-0.189	-0.447	
45-49	-0.018	-0.062	-0.020	-0.101	-0.112	-0.213	
50-54	-0.020	-0.058	-0.068	-0.146	-0.060	-0.206	
55-59	-0.011	-0.017	-0.064	-0.093	0.005	-0.088	
60-64	-0.003	-0.003	-0.001	-0.007	-0.001	-0.008	
Total	-0.186	-0.748	-0.294	-1.228	-1.187	-2.425	

Sources: Computed from the original Goskomstat's tables of deaths and population by age and educational status.

## 6 Variations in life expectancy and Gini coefficient in time and across countries

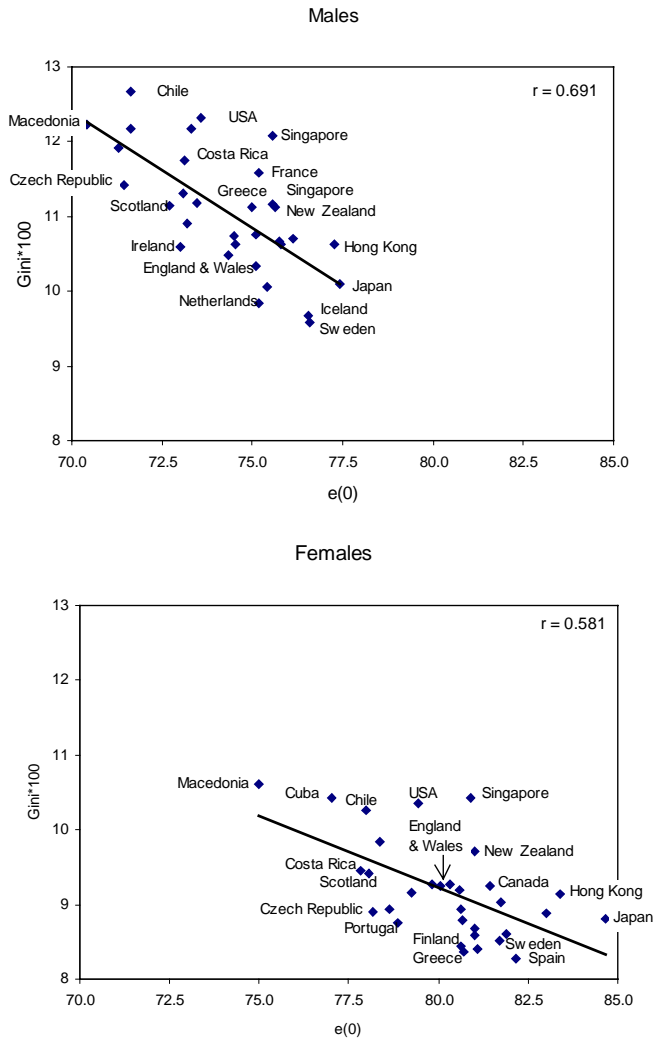
Prior studies of historical trends in inter-individual inequality in length of life have proved two fundamental facts about changes in the mortality of human populations (Illsey and Le Grand, 1987, Llorca et al., 1998, Wilmoth and Horiuchi, 1999, Anand and Nanthikesan, 2001). First, during the 20th century the inter-individual inequality (or variability) in length of life had been declining, mirroring the increase in average length of life. Second, during the last three decades, this correlation has become weaker since life expectancy has continued to increase, while the decline in the inter-individual inequality in length of life has slowed down or even stopped in low mortality countries. Both facts can be observed in all countries having a long series of mortality statistics.

Wilmoth and Horiuchi (1999) explain this regularity by a principal historical change in the age pattern of mortality. The historical lowering of mortality rates was much more pronounced in the young than in old ages. Therefore, life table deaths have been more and more concentrated at old ages. After a certain point (in the 1950s, 1960s or 1970s, depending on the country) at which mortality at young ages had already been reduced to low values, its further reduction was unable to reduce significantly dispersion of ages at death. In addition, in the 1980s-1990s, the mortality decline in countries with low mortality was more pronounced at old ages than at middle ages. This process (as it was shown in our earlier examples) increases inequality in length of life.

This means that old-age deaths are still partly balanced by a considerable proportion of deaths at ages which are considerably younger than the average length of life. In some countries with a relatively high average level of length of life, young and middle-age deaths are not as low as they could be.

Our experiments (not shown here) with the values of Gini coefficient and life expectancy for about 45 countries, for the period of 1960s-1990s, suggest the following. If one considers a wide variety of populations with very different life expectancies, then the negative association between life expectancy at birth and Gini coefficient for the full range of ages would be very tight. The correlation coefficients vary from -0.85 to -0.95, depending on the selection of countries and years. If only countries with comparable levels of mortality are selected, then this correlation is substantially weaker.

Figure 6 displays the positions of 32 countries according to male life expectancy at birth, and Gini coefficient (full range of ages). In all these countries male life expectancy at birth was higher than 70 years in 1994-99. The correlation coefficient between life expectancy and Gini coefficient by country is -0.69 for men and -0.58 for women. In many cases, the same, or almost the same life expectancies correspond to different levels of Gini coefficient. For example, in the USA the male  $e_0$  is 73.6 with

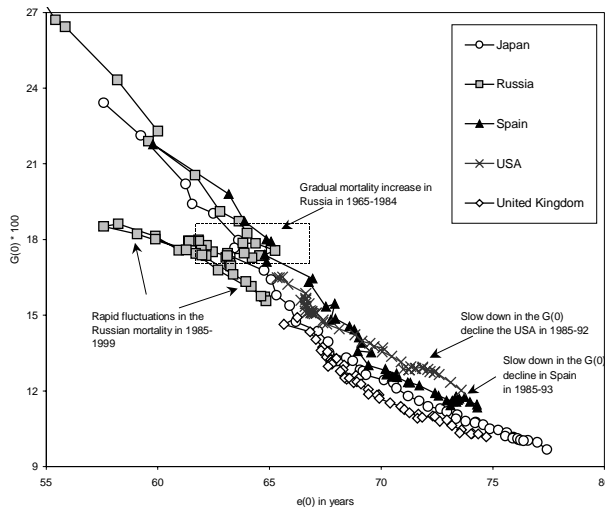


**Figure 5:** Relationship between life expectancy and Gini coefficient in 1996-99 for men and women in 31 countries with male life expectancies higher or equal to 70 years.

Sources: Data for computations are extracted from the WHO Mortality Database (2001)

$G_0 \cdot 100$  12.3, while in Ireland, the equivalent figures are 73.0 and 10.6. In Chile, the USA, Cuba and Singapore, the values of  $G_0 \cdot 100$  for the male population are substantially higher than those predicted by life expectancy. On the other hand, in the Czech Republic, Ireland, the Netherlands, Norway and Sweden, they are lower. For women, Chile, the USA and Singapore have an "excess" in Gini coefficient, while the Czech Republic, Portugal, Greece and Sweden have comparatively low values of Gini coefficient.

Trajectories of the male populations of five countries (Japan, Russia, Spain, USA, UK) in 1950-99 in the coordinates  $e_0$  (horizontal axis) and  $G_0 \cdot 100$  (vertical axis) are shown in Figure 6. The countries started in 1950 from very different levels of life expectancy and Gini coefficient. In Japan, the values of the two indicators were 58 and 23, in Russia they were 52 and 31, in Spain they were 59 and 22, in the USA they were 65 and 17, and in the UK they were 66 and 15. Since then, all the countries, except Russia, have experienced a continuous progression in the lengthening of life and a reduction of inequality in length of life.



**Figure 6:** Trajectories in coordinates  $e_0$  and  $G_0 \cdot 100$  for the male populations of Japan, Russia, Spain, USA, and the UK in 1950-99.

Sources: Data for computation are extracted from the WHO Mortality Database (2001)

The case of Russia is remarkable. After impressive improvements in the 1950s, the mortality of men had been increasing since the mid-1960s. The chaotic fluctuations of the points for Russia correspond to a rapid increase in life expectancy in the second half of the 1980s, after Gorbachev's anti-alcohol campaign, and to its striking fall in the first half of the 1990s (Shkolnikov et al., 1996).

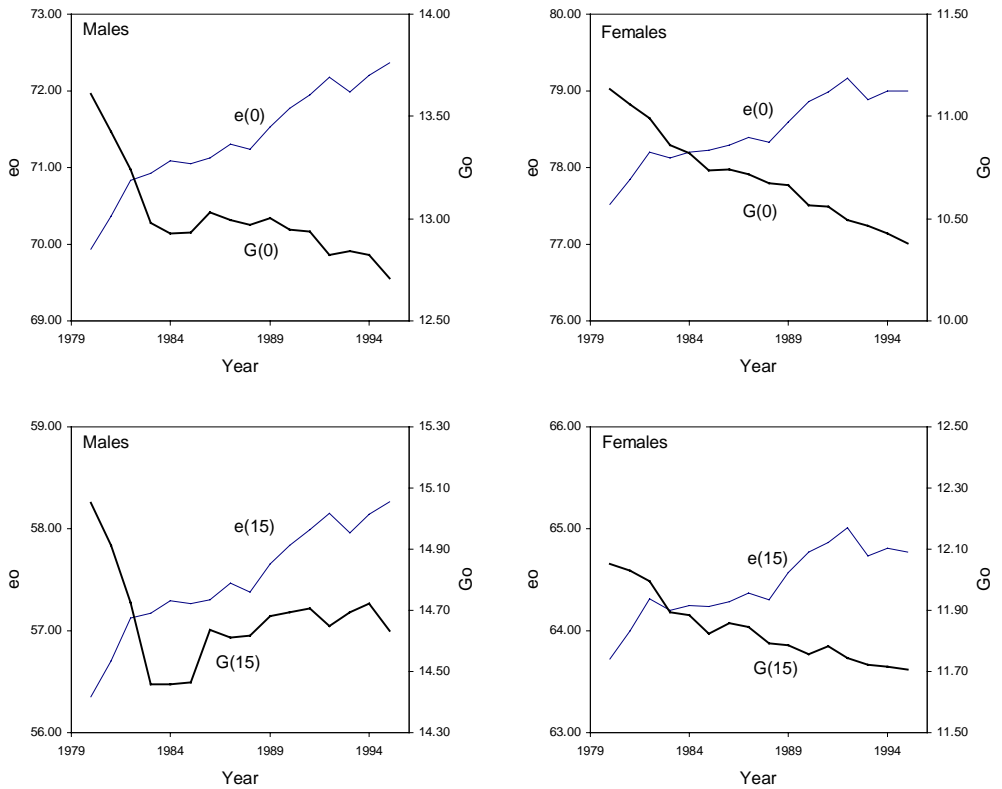
Japan and Spain experienced a spectacular increase in the average length of life and also in the reduction of inequality in length of life starting from very low levels of length of life and high inequality.  $G_0$  in Japan has always been lower than that in the USA after the level of 65 years of male life expectancy. In Spain,  $G_0$  has become lower than that in the USA, starting from the points where life expectancy of men becomes higher than 69 years.

The USA and the UK are two countries where starting levels of mortality in the 1950s were much lower than those in Japan, Russia or Spain. Male life expectancy in these countries has been rather similar with only one year in favor of the UK during almost the whole period of observation. There is, however, a considerable difference between the two countries in  $G_0$ , which appears to be unusually low in the UK (see also Figure 3).

Finally, one can see periods of flattening in Gini coefficient for males in the USA and Spain in the 1980s and early 1990s (Figure 6).

Figure 7 represents a zoomed image of changes in the male and female life expectancies and the Gini coefficients in the USA in 1980-1996. For men, the life expectancy at birth continues increasing over this period, while the Gini coefficient does not experience any decline between 1983 and 1990 (upper-left panel). If we exclude the influence of infant mortality from consideration, then the pattern becomes clearer for men because the life expectancy at age 15 increases, while the Gini coefficient for ages above 15 increases as well, between 1983 and 1994 (lower left panel). The difference between men and women is very clear because for women the situation appears normal, both for the full range of ages and also for ages 15+: life expectancy increases and the Gini coefficient declines.

Decomposition of the increase in  $G_{15} \cdot 100$  for US men between 1983 (beginning of the increase) and 1991 (a point of maximum) shows that it can be explained by two major factors (Table 6). First, deterioration in mortality of adults at ages from 15 to 44 mainly due to AIDS and also due to violence at ages from 15 to 29. Second, a further decrease in mortality at old ages. The first factor is clearly related to contemporary health hazards and social environment, whereas the second one reflects the influence of the general evolution of the overall shape of distribution of deaths by age.



**Figure 7:** Trends in life expectancy and Gini coefficient (ages 0 and 15) for men and women in the USA in 1980-95.

Sources: Data for computations are extracted from The Berkeley Mortality Database (2001). Life tables for the United States are based on those constructed by the Office of the Actuary of the Social Security Administration.



In Spain the nature of the increase in Gini coefficient since 1985 is similar to that in the USA. Llorka and colleagues (1998) explain the increase in the Gini coefficient after 1985, mostly by rising mortality from AIDS. Our decomposition (not shown here) of these changes by age and cause of death agrees with this finding.

**Table 6:** *Components of the increase in  $G_{15} \cdot 100$  for US men between 1983 and 1991\* by age and cause of death*

Age group	All causes combined	Infectious diseases	Cancers	Endocrine, metabolic, and immunity disorders including AIDS	Diseases of the circulatory system	Diseases of the respiratory system	Diseases of the digestive system	Other diseases	Accidents and violence
15-29	0.133	0.004	-0.008	0.088	-0.005	0.002	-0.003	0.002	0.053
30-44	0.202	0.015	-0.023	0.268	-0.063	0.009	-0.005	0.028	-0.028
45-59	-0.191	0.005	-0.030	0.074	-0.193	-0.002	-0.021	-0.004	-0.020
60-74	-0.041	0.000	0.000	0.004	-0.036	-0.001	-0.004	-0.003	-0.002
75+	0.193	-0.007	-0.018	-0.010	0.288	-0.037	0.002	-0.027	0.001
Total 15+	0.296	0.016	-0.078	0.424	-0.008	-0.030	-0.031	-0.002	0.005

\*  $G_{15} \cdot 100$  was 14.6 in 1983 and 14.9 in 1991

Sources: Data for computation are extracted from the WHO Mortality Database (2001)

## 7 Conclusion

The key results of this paper can be summarized as follows:

1. Gini coefficient and the other four common measures of inequality (the interquartile range, the standard deviation, Theil entropy index, and the variance of the logarithm of length of life) are defined in continuous and discrete forms for length-of-life distributions. Gini coefficient satisfies three basic properties, desirable for inequality measures, and correctly reflects relations of the Lorenz-dominance among length-of-life distributions. This makes a difference from the standard deviation and the interquartile range.

2. Consideration of temporal changes in Gini coefficient and the four other inequality measures in the USA and Russia since the 1950s suggests that the choice of inequality measure can matter for the judgements about the direction of changes in the inequality. Gini coefficient, unlike some other inequality measures, is not extremely

sensitive to changes in infant mortality and reasonably sensitive to changes at adult ages.

3. The toolkit for calculation of Gini coefficient from the life table data and various types of decompositions of difference between two Gini coefficients are proposed.

4. A method for calculation of Gini coefficient from the life table data is based on simple formulae for numerical integration of function  $[l(x)]^2$  within elementary age intervals, including special formulae for age 0 and the open-ended age interval 85+. The method has been tested on more than 400 real life tables. It gives an opportunity for estimating Gini coefficient from abridged mortality data with age groups 0, 1-4, 5-9, ..., 85+, which is a typical format for the WHO mortality data.

5. New formulae for the decomposition of differences between two Gini coefficients, by age and cause of death, have been developed, using the general replacement algorithm. Empirical examples reveal meaningful differences between components of temporal changes, or variations, in Gini coefficient, from the components of changes or variations in life expectancy due to greater effects of changing mortality at young ages, and opposite effects of mortality at advanced ages. Effects of elimination of causes of death on Gini coefficient depend on age distribution of deaths from these causes. In the UK in the 1990s, elimination of cardiovascular disease and respiratory disease would lead to an increase in Gini coefficient, while elimination of external causes, infectious disease and cancer, would lead to a decrease.

6. A method for decomposition of differences between two Gini coefficients, in respect to age and population group, has been developed, by applying the replacement algorithm. This type of decomposition allows linking inter-individual variability in length of life with inter-group differences in length of life and the population weights of groups. In Russia only half of the overall decrease in Gini coefficient between 1979 and 1989, for men at ages from 20 to 64, is due to changes in mortality, whereas another half was produced by an improvement of the educational composition of the population.

7. Analyses of variation in life expectancy and Gini coefficient, in time and across countries, has been performed. Cross-sectional comparisons of countries show a strong negative correlation between both measures and some variations in Gini coefficient at the same level of life expectancy. Periods of increase or stagnation in Gini coefficient, with simultaneously increasing life expectancies, were detected in the  $G_0-e_0$  trajectories for the USA and Spain in the 1980s and the early 1990s. They are associated with increases in mortality of young adults due to HIV/AIDS and violence. Degree of inequality in length is determined by evolution of mortality curves and depends on specific epidemiological and social patterns.

## **8 Acknowledgements**

Early stages of the study were completed in 1997-99, as a part of the Global Health Equity Initiative of the Rockefeller Foundation.

We are grateful to Professor Sudhir Anand, who motivated our interest in the subject and with whom some of ideas were discussed in 1997-98. We are also grateful to Professor John Wilmoth and his team at the University of California at Berkeley for their work on The Berkeley Mortality Database. This rich data source on mortality trends and age patterns has been extensively used in our study.

We are grateful to S.Nanthikesan, T.Evans, F.Diderichsen, and M.Whitehead for the opportunity to discuss various issues related to inequalities in health and its measurement, in 1997-98.

Finally, we are grateful to the two anonymous reviewers of *Demographic Research* for their helpful and constructive comments on the manuscript.

## Notes

1. Material of this section is adapted from Anand (1983).
2. Murrey, Frenk and Gakidou (2001) proposed a Gini-shape index combining absolute and relative inter-individual inequalities.
3. *STD* and *VAR* as well as  $G_0^{abs}$  are invariant to a shift in individual ages at death. These measures do not change if all individual ages of death increase or decrease by the same number of years. It is not clear, however, how such shifts can happen in reality. When each age at death increases by the same amount of time  $t$  (a rightward shift) then nobody will die at ages from 0 to  $t$ . A similar leftward shift will result in negative ages for those dying at ages 0 to  $t$ .
4.  $A_{85+}$  is defined here as the average duration of life after age 85 and is measured in years. This makes a difference from parameters  $A_x$  for closed age intervals.
5. Preliminary analysis (not shown here) suggested that a linear relationship fits well the empirical data.

## References

- Allison, P.D. 1978. Measures of inequality. *American Sociological Review* 43, p. 865-880.
- Anand, S. 1983. *Inequality and Poverty in Malaysia: Measurement and Decomposition*. New York: Oxford University Press.
- Anand, S. and Nanthikesan, S. 2000. A compilation of length-of-life distribution measures for complete life tables. *Working Papers Series*, 10(7), Harvard Center for Population and Development Studies, Cambridge, MA, June.
- Anand, S. and Nanthikesan, S. 2000. The evolution of inequality in length of life: France in 1806-1997. *Working Papers Series*, 11(3), Harvard Center for Population and Development Studies, Cambridge, MA, February.
- Anand, S., Diderichsen, F., Evans T., Shkolnikov V. and Wirth, M. 2001. Measuring disparities in health: methods and indicators. In: *Challenging Inequities in Health: From Ethics to Action*. ed. by T.Evans, M.Whitehead, F.Diderichsen, A. Bhuiya. New York: Oxford University Press: 49-67
- Andreev, E.M. 1982. Metod komponent v analize prodoljitelnosti zjizni. [The method of components in the analysis of length of life]. *Vestnik Statistiki*, 9: 42-47.
- Andreev, E.M., Shkolnikov, V.M. and Begun, A.Z. 2002. Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. *Demographic Research*, vol. 7, article 14.
- Arriaga, E. 1984. Measuring and explaining the change in life expectancies. *Demography* 21(1): 83-96.
- Atkinson, A.B. 1970. On the measurement of inequality. *Journal of Economic Theory*, 2: 244-63.
- Atkinson, A.B. (ed). 1980. *Wealth, Income and Inequality*, 2nd edn, Oxford: Oxford University Press.
- Bourgeois-Pichat, J. 1951. La mesure de la mortalite infantile. Principes et methodes. *Population*, 2: 233-248.
- Das Gupta, P. 1994. Standartization and decomposition of rates from cross-classified data. *Genus*, L(3-4): 171-196.

- Das Gupta, P. 1999. Decomposing the difference between rates when the rate is a function of factors that are not cross-classified. *Genus*, LV(1-2): 9-26.
- Fries, J.F. 1980. Aging, natural death and the compression of morbidity. *New England Journal of Medicine*: 303(3): 130-135
- Goodwin, D.G. and Vaupel, J.W. 1985. Curves and have-statistics for ecological analysis of diversity. Part III: Comparison of measures of diversity. Working Paper WP-85-91, International Institute for Applied System Analysis, Laxenburg, Austria..
- Hanada, K. 1983. A formula of Gini's concentration ratio and its application to life tables. *Journal of Japan Statistical Society*, 13: 95-98.
- Hicks, D.A. 1997. The inequality-adjusted human development index: a constructive proposal. *World Development*, 25(8): 1283-1298.
- Horiuchi, S., Wilmoth J.R., and Pletcher, S.D. 2001. A general method of decomposition analysis. *Unpublished manuscript*.
- Human Life-Table Database. 2002. <http://www.lifetable.de>. Last visited in October 2002.
- Illsey, R. and Le Grand, J. 1987. The measurement of inequality in health. In: *Health and Economics*. Ed. by A.Williams. London: Macmillan: 13-36.
- Interstate Statistical Committee of CIS. 1996. *Results of the 1989 All-Union Population Census*. CD-ROM.
- Kannisto, V., Lauritsen, A., Thatcher, R. and Vaupel, J.W. 1994. Reduction of mortality at advanced ages. *Population and Development Review* 20: 793-810.
- Kendall, M.G., Stuart, A. 1966. *The Advanced Theory of Statistics*. London: Charles Griffin.
- Kitagawa, E. 1964. Standardized comparisons in population research. *Demography*, 1: 296-315.
- Leclerc, A., Lert, F. and Fabien, C. 1990. Differential mortality: some comparisons between England and Wales, Finland and France, based on inequality measures. *International Journal of Epidemiology* 19: 1001-1010.
- Le Grand, J. 1987. Inequality in health: some international comparisons. *European Economic Review*, 31: 182-192.

- Le Grand, J. 1989. An international comparison of distributions of ages-at-death. In: Fox, J. (ed), *Health Inequalities in European Countries*. Aldershot: Gower Press.
- Llorca, J., Prieto, D.M., Alvarez, C.F., Delgado-Rodriguez, M. 1998. Age differential mortality in Spain, 1990-1991. *Journal of Epidemiology and Community Health*, 52: 259-261.
- Lynch, S. and Brown, S. 2001. Reconsidering mortality compression and deceleration: an alternative model of mortality rates. *Demography*, 38(1): 79-95.
- Matlab Report. 1996. *Demographic Surveillance System – Matlab. Registration of Vital Events*. Vol. 27, Scientific Report No 76, Dhaka, December.
- Murray, C.J.L., Frenk, J. and Gakidou, E.E. 2001. Measuring health inequality: challenges and new directions. In: D.Leon and G.Walt eds. *Poverty, Inequality and Health*. Oxford University Press: 194-216.
- Myers, G.C. and Manton, K.G. 1984. Compression of mortality: myth or reality? *The Gerontologist*, 24(4): 346-353.
- Pollard, J.H. 1982. The expectation of life and its relationship to mortality. *Journal of the Institute of Actuaries*, 109 (442), Part II: 225-240.
- Pollard, J.H. 1988. On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography*, 25: 265-276.
- Pressat, R. 1985. Contribution des écarts de mortalité par âge à la différence des vies moyennes. *Population*, 4-5: 766-770.
- Preston, S.H., Heuveline, P., and Guillot, M. 2001. *Demography. Measuring and Modelling Population Process*. Blackwell Publishers. Oxford, Malden, Massachusetts.
- Sen, Amartia. 1973. *On economic inequality*. Oxford: Clarendon Press.
- Shkolnikov, V.M., Leon, D.A., Adamets, S., Andreev, E., Deev, A. 1998. Educational level and adult mortality in Russia: an analysis of routine data 1979 to 1994. *Social Science and Medicine*, 47(3): 357-369.
- Shkolnikov, V.M., Meslé, F. and Vallin, J. 1996. Health crisis in Russia. *Population: An English Selection*, 8: 123-190.
- Shkolnikov, V., Valkonen, T., Begun, A. and Andreev, E. 2001. Measuring inter-group inequalities in length of life. *Genus*, LVII(3-4): 33-62.

- Silber, J. 1988. On inequality before death and life table summary measures. *Genus*, XLVI:25-39.
- Silber, J. 1992. Inequality in mortality: measuring the contributions of various causes of death. *Genus*, Vol. XLVIII, No 3-4, pp. 95-98.
- The Berkeley Mortality Database. 2001. <http://www.demog.berkeley.edu/wilmoth/mortality>. Last visited in June 2001.
- Theil, H. 1967. *Economics and information theory*. North-Holland, Amsterdam.
- Wilmoth, J.R. and Lundström, H. 1996. Extreme longevity in five countries: presentation of trends with special attention to issues of data quality. *European Journal of Population*, 12: 63-93.
- Wilmoth, J.R. and Horiuchi, S. 1999. Rectangularization revisited: variability in age at death within human populations. *Demography*, 36( 4): 475-495.
- WHO Mortality Database. 2001. <http://www.who.int/whosis/mort/download.htm>  
Last visited in August 2002.



## Appendix 1

According to Wilmoth and Horiuchi (1999), the interquartile range *IQR* is the difference between the 75% and the 25% quartiles of the  $l(x)$  distribution. It is defined as

$$IQR = x_{25} - x_{75},$$

where  $x_{25}$  and  $x_{75}$  are such that  $l(x_{25}) = 0.25, l(x_{75}) = 0.75$ , (A)

with the radix  $l(0)=1$ .

The variance *VAR* and the standard deviation *STD* are

$$VAR = \int_0^{\infty} (x - \mu)^2 f(x) dx, \text{ where } \mu = e(0), f(x) = d(x)/l(0) = d(x), \quad (B)$$

$$STD = \sqrt{VAR} \quad . \quad (C)$$

The variance of the logarithm of length of life *VarLog* is

$$VarLog = \int_0^{\infty} (\log x - \log \hat{\mu})^2 f(x) dx, \quad (D)$$

where the geometric mean  $\hat{\mu}$  is determined by  $\log \hat{\mu} = \int_0^{\infty} f(x) \log x dx$ .

The Theil entropy index is

$$T = \int_0^{\infty} \left( \frac{x}{\mu} \log \frac{x}{\mu} \right) f(x) dx. \quad (E)$$

Taking into account that  $d(x) = -\frac{dl(x)}{dx}$  equations (B) and (E) can be transformed in

$$VAR = 2 \int_0^{\infty} xl(x) dx - \mu^2, \quad (B1)$$

$$T = \frac{1}{\mu} \cdot \int_0^{\infty} l(x) \log x dx + (1 - \log \mu) \quad (E1)$$

by taking integrals in (B) and (C) by parts.

Calculation of *IQR* from the life table data is easy. Values  $x_{25}$  and  $x_{75}$  can be interpolated from the corresponding  $l_x$  values. The estimates would be less precise if calculated from abridged life tables. For low-mortality populations it is important to have life table data for advanced ages since the age  $x_{25}$  often exceeds 85 years.

It is more difficult to find the best way for a practical computation of other inequality measures. For each of the measures one has to develop and test an appropriate algorithm (as it is done for the Gini coefficient in section 3). Although this task is far beyond the scope of the present paper, we will briefly indicate some possibilities.

In studies by Anand and colleagues (Anand et al., 2001, Anand and Nanthikesan, 2000) and also in section 5 of the present study, a straightforward approach was taken, wherein *VAR* is estimated as

$$VAR \cong \sum_{x=0}^{\omega} d_x \cdot (\bar{x} - e_0)^2, \tag{B2}$$

where  $\bar{x}$  is the average age at death within the elementary age interval  $[x, x+1)$ . Similarly, *VarLog* is estimated as

$$VarLog \cong \sum_{x=0}^{\omega} d_x \cdot (\log \bar{x} - \log \hat{\mu})^2, \text{ where } \log \hat{\mu} \cong \sum_{x=0}^{\omega} d_x \cdot \log \bar{x} \tag{D2}$$

and *T* is estimated as

$$T \cong \frac{1}{e_0} \sum_{x=0}^{\omega} (d_x \cdot \bar{x} \cdot \log \frac{\bar{x}}{e_0}). \tag{E2}$$

Function  $d(x)$  changes within elementary age intervals rather quickly, especially at ages around its mode. It suggests that estimates (B2)-(E2) could be imprecise for abridged life tables.

One might prefer to use formulae (B1) and (E1) with a more stable cumulative function  $l(x)$  for numerical integration. Relevance of different approaches to numerical integration and precision of respective estimates for different types of life tables deserve an additional detailed investigation.

## Appendix 2

Raising to square of right and left parts of the equation (5) gives

$$[l(x+t)]^2 = (l_x)^2 + 2l_x(l_{x+1} - l_x)t + 12C_x l_x(l_{x+1} - l_x)(t^2 - t) + (l_{x+1} - l_x)^2 [36(C_x)^2 t^4 + (12C_x - 72(C_x)^2)t^3 + (1 - 12C_x + 36(C_x)^2)t^2]$$

The integral of this function over  $t$  from 0 to 1 is

$$\int_0^1 [l(x+t)]^2 dt = l_x l_{x+1} - 2C_x l_x (l_{x+1} - l_x) + \left[ \frac{1}{3} - C_x - \frac{6}{5}(C_x)^2 \right] \cdot (l_{x+1} - l_x)^2.$$

This expression, together with (4), allows for deriving  $\hat{A}_x$  after simple transformations

$$\hat{A}_x = \frac{1 - \frac{2}{3}q_x + C_x(2 - q_x - \frac{6}{5}C_x)}{2 - q_x}.$$

### Appendix 3

Within the first year of life, the survival function can be described by the following "law" (Borgeois-Pichat, 1951)

$$l(t) = 1 - \alpha - \beta \ln^3(365 \cdot t + 1), \quad 0 \leq t \leq 1$$

Parameters  $\alpha$  and  $\beta$  can be derived from two constrains

$$l_1 = 1 - \alpha - \beta \cdot \ln^3(366) = 1 - \alpha - \beta \cdot C. \quad (\text{A})$$

$$L_0 = 1 - \alpha - \beta \int_0^1 \ln^3(365 \cdot t + 1) dt = 1 - \alpha - \beta \cdot S_3 \quad (\text{B})$$

The constants  $C$  and  $S_3$  are universal ones. They do not depend on the population under study and should be calculated only once. Since  $L_0 = l_1 + A_0(1 - l_1)$ , equations (A) and (B) yield

$$\begin{aligned} \alpha &= q_0 - \beta \cdot C, \\ \beta &= \frac{A_0 q_0}{C - S_3} \end{aligned} \quad (\text{C})$$

$[l(t)]^2$  is obviously equal to  $(1 - \alpha)^2 - 2(1 - \alpha)\beta \ln^3(365 \cdot t + 1) + \beta^2 \ln^6(365 \cdot t + 1)$ .

Its integral over  $t$  from 0 to 1 is

$$\int_0^1 [l(t)]^2 dt = (1 - \alpha)^2 - 2(1 - \alpha)\beta \cdot S_3 + \beta^2 \cdot S_6, \quad (\text{D})$$

where  $S_6 = \int_0^1 \ln^6(365 \cdot t + 1) dt$ .

From definition (4), it follows that  $\hat{A}_0 = \frac{\int_0^1 l(x) dx - (l_1)^2}{1 - (l_1)^2}$ . Using this expression and (D), values of the constants  $\ln(366)=5.9026$ ,  $C=205.6541$ ,  $S_3=130.9208$ ,

$S_6=20106.7400$  and equations (C) for parameters  $\alpha$  and  $\beta$  leads to the final formula for  $\hat{A}_0$

$$\hat{A}_0 = A_0 \left( 1 - q_0 \frac{3 + 0.83087a_0}{2 + q_0} \right)$$

## Appendix 4

The difference between two values  $G_0^{t\text{abs}}$  and  $G_0^{\text{abs}}$  produced by mortality replacement at age  $x$  and younger ages can be expressed as

$$\mathcal{E}_{0,x}^{\text{abs}} = G_0^{\text{abs}}(M^{[x]}) - G_0^{\text{abs}} = e_0(M^{[x]}) - \int_0^{\infty} [l'_t(M^{[x]})]^2 dt - e_0 + \int_0^{\infty} [l'_t]^2 dt = e_0(M^{[x]}) - e_0 + \theta_0 - \theta'_{0|x} - \theta_x(l'_x)^2.$$

The decomposition of the difference between two absolute Gini measures by elementary age group, similar to (14) and (17), is

$$G_0^{t\text{abs}} - G_0^{\text{abs}} = \sum_{i=0}^{n-1} (\mathcal{E}_{0,x_{i+1}}^{\text{abs}} - \mathcal{E}_{0,x_i}^{\text{abs}}) = \sum_{i=0}^{n-1} \mathcal{E}_i^{\text{abs}}, \text{ where}$$

$$\mathcal{E}_i^{\text{abs}} = G_0^{\text{abs}}(M^{[x_{i+1}]}) - G_0^{\text{abs}}(M^{[x_i]}).$$

Similarly to (20), for a small  $\Delta x$  we get

$$\lambda_x^{\text{abs}} = \mathcal{E}_{x,x+\Delta x}^{\text{abs}} / \Delta x = (\mu_x - \mu'_x) l'_x (e_x - 2l'_x \theta_x) = (\mu_x - \mu'_x) \cdot \eta_x^{\text{abs}}.$$

If there are  $m$  causes of death, then for each age group

$$\mathcal{E}_{x_i, x_{i+1}}^{\text{abs}} \cong \left( \int_{x_i}^{x_{i+1}} \eta_t^{\text{abs}} dt \right) \cdot \sum_{j=1}^m (m_{x_i|x_{i+1},j} - m'_{x_i|x_{i+1},j}).$$

If in each age group  $[x_i, x_{i+1})$   $m_{x_i|x_{i+1}} \neq m'_{x_i|x_{i+1}}$  then cause-specific components within this age group are simply

$$\mathcal{E}_{x_i, x_{i+1}, j}^{\text{abs}} = \frac{m_{x_i|x_{i+1},j} - m'_{x_i|x_{i+1},j}}{m_{x_i|x_{i+1}} - m'_{x_i|x_{i+1}}} \cdot \mathcal{E}_{x_i, x_{i+1}}^{\text{abs}},$$

Otherwise, a numerical integration would be necessary to compute cause-specific components according to

$$\mathcal{E}^{abs}_{x_i, x_{i+1}, j} \cong \left( \int_{x_i}^{x_{i+1}} \eta^{abs}_t dt \right) \cdot (m_{x_i | x_{i+1}, j} - m'_{x_i | x_{i+1}, j}).$$

