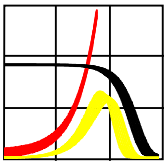


Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Doberaner Strasse 114 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
http://www.demogr.mpg.de

MPIDR WORKING PAPER WP 2001-017
JUNE 2001 (REVISED SEPTEMBER 2004)

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

Vladimir M. Shkolnikov (shkolnikov@demogr.mpg.de)
Evgeniy E. Andreev (evg_andreev@ns.cnt.ru)
Alexander Z. Begun (begun@demogr.mpg.de)

*Draft 3. 26 June 2001*

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

Vladimir M. Shkolnikov

Max Planck Institute for Demographic Research, Rostock
Shkolnikov@demogr.mpg.de

Evgeniy E. Andreev

Centre of Demography and Human Ecology, Moscow
evg_andreev@ns.cnt.ru

Alexander Z. Begun

Max Planck Institute for Demographic Research, Rostock
Begun@demogr.mpg.de

Abstract

This paper presents a toolkit for measuring and analysing inter-individual inequality in length of life by Gini coefficient. Gini coefficient is treated as an additional function of the life table. A new method for the estimation of Gini coefficient from life table data has been developed and tested on the basis of hundreds of life tables. The method provides precise estimates of Gini coefficient for abridged life tables even if the last age group is 85+. New formulae have been derived for the decomposition of differences in Gini coefficient by age and cause of death. A method for further decomposition of age-components into effects of mortality and population group has been developed. It permits the linking of inter-individual inequalities in length of life with inter-group inequalities. Empirical examples include the decomposition of secular decrease in Gini coefficient in the USA by age, decomposition of the difference in Gini coefficient between the UK and the USA by age and cause of death, temporal changes in the effects of elimination of causes of death on Gini coefficient, and decomposition of changes in Gini coefficient in Russia by age and educational group. Consideration of the variations in Gini coefficient during the last decades and across modern populations show that these variations are driven not only by historical shifts in the distribution of deaths by age, but also by peculiar health and social situations.

Send correspondence to the first author by e-mail: Shkolnikov@demogr.mpg.de

Running title: Gini coefficient as a life table function

Keywords: inequality, life expectancy, Gini coefficient, mortality

## *Introduction*

In 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 heterogeneity in respect to length of life attract growing attention (Anand et al., 2001). Gini coefficient is a very useful measure of inequality in age at death (or length of life) in practical terms reflecting degrees of variability in age at death. As a formal construct Gini coefficient has significant similarities with life expectancy, which allow applying similar methods for its calculation or decomposition.

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 or wealth (Atkinson, 1970 and 1980, Sen, 1973, Anand, 1983).

In a number of studies Gini coefficient has been applied to mortality schedules. It has been employed for analyses of the variation in degree of people's inequality in the face of death over time and across countries. In some studies, Gini coefficient was 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, 1989, 1992, Llorka et al., 1998).

Raimond Illsey and Julian 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 at working ages or do not work for some other reasons.

By all means, it is worth understanding that greater or lower values of Gini coefficient or other measures of individual variability in age at death show a greater or lower magnitude of inter-individual differences, but do not necessarily correspond to greater or lower degrees of social inequality in mortality/health. A principal divide between group-based and individual-based ways of measuring inequalities played a major role in the dispute of the results of the famous Black's report (Black et al., 1980) by Le Grand and Rabin (1986).

Illsey and Le Grand (1987) computed Gini coefficient from distributions of deaths by age in real populations. In order to avoid a bias due to different age structures, a standard population age structure was used for weighting. This approach is the same as that in economics since age at death and population distributions are independent from each other, exactly like income and population in econometrics.

Other researchers have 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, Nanthikesan, 2000, Anand et al., 2001). In this case, the constitution of inequality indices becomes more complex since "population" is determined by the life table cohort's survival and, therefore, depends on mortality.

Gini coefficient has also been considered, among other indices of inequality, as a possible measure of the rectangularization of survival curves in human populations (Wilmoth and Horiuchi, 1999). This approach is closely linked to the wider agenda 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 present study has a more utilitarian purpose. It provides a toolkit for operating with Gini coefficient as a possible standard function of the life table. Using some of its features we have introduced a simple method for its computation from discrete data. We have also given a new formula for decomposing differences between two Gini coefficients by age and cause of death. We have applied a more general algorithm for additional decomposition of differences in Gini coefficient or life expectancy by population group. Finally, variations in Gini coefficient in time and across countries during the last decades have been considered.

## *1. Definitions*

The most common geometric definition of Gini coefficient is based on the Lorenz (or concentration) curve. It represents cumulative income share as a function of cumulative population share. If a population share is always exactly equal to a share in overall income then there is a situation of perfect equality.

Applying this device 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". Hence, 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 $\Phi_x = \frac{\sum_{t=0}^{x-1} d_t \cdot \bar{t}}{\sum_{t=0}^{\omega-1} d_t \cdot \bar{t}} = \frac{T_0 - T_x - l_x \cdot x}{T_0}$, where $\omega$ 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*.

For each Lorenz curve Gini coefficient is defined as an area between the diagonal and this curve, divided by the whole area below the diagonal (equal to 1/2). If one believes that the Lorenz curve can be approximated by a number of straight line segments, then Gini coefficient can be computed as (Anand, 1983):

$$G_0 = 1 - \sum_{x=0}^{\omega-1} (F_{x+1} - F_x)(\Phi_{x+1} + \Phi_x) \qquad (1)$$

--- Figure 1 about here ---

Figure 1 shows Lorenz curves for three female life tables describing the very different female mortality patterns of Sweden, Russia, and Bangladesh in 1995. The higher the level of life expectancy at birth, the lower the level of inequality in age at death.

In formula (1) the sum covers the whole range of ages. Apparently, the upper and lower limits of summation in formula (1) can be changed. For example, one might be interested in the inequality in age at death for ages above 15 ($G_{15}$) or between 20 and 65 ($G_{20|65}$). The values of Gini

coefficient can be computed for different ranges of ages exactly as for life expectancy. Figure 2 demonstrates $1-G_x$ curves for Swedish men and women in 1920 and 1995.

----- Figure 2 about here ---

The definition of the Gini coefficient by Kendall and Stuart (1966) is also helpful in understanding its nature. If the population under consideration consists of $n$ 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

$$G_0 = \frac{1}{2n^2 e_0} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| x_i - x_j \right|.$$

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

$$G_0 = \frac{1}{2l_0{}^2 e_0} \sum_{x=0}^{\omega} \sum_{y=0}^{\omega} d_x \cdot d_y \cdot \left| \bar{x} - \bar{y} \right|. \qquad (2)$$

Gini coefficient has a number of properties described elsewhere (Anand, 1983, Goodwin and Vaupel, 1985). These are, for example, mean and scale independence and population size independence. The coefficient varies between 0 and 1. It is equal to zero if all people in a population die at the same age, and equal to 1 if all people die at age 0 and one individual at an infinitely old age.

Discrete approximation (1) of the Gini coefficient is satisfactory only if deaths are evenly or almost evenly distributed within each elementary age interval. This condition can be violated in the case of 5-year age intervals in abridged life tables. Formula (2) is simple to understand and exact, but hardly applicable for practical calculations.

For further consideration we will use a continuous expression of the Gini coefficient derived by K.Hanada (1983)

$$G_0 = 1 - \frac{1}{e_0} \cdot \int_0^{\infty} [l(x)]^2 \, dx \qquad (3)$$

Most importantly, this formula is quite similar to the one for life expectancy $e_0 = \int_0^{\infty} l(x)dx$. It suggests that for practical calculation the area under the curve $[l(x)]^2$ should be estimated similarly to the area under curve $l(x)$ for the life expectancy.

## 2. Computation from discrete data: 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 $\int_0^{\infty} [l(x)]^2 dx$.

In a complete life table the life expectancy at birth is estimated as

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

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 $A_x = \dfrac{(L_x/1) - l_{x+1}}{l_x - l_{x+1}}$.

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

$$\sum_x \int_0^1 [l(x+t)]^2\,dt = \sum_x \left[(l_{x+1})^2 + \hat{A}_x((l_x)^2 - (l_{x+1})^2)\right]. \qquad (4)$$

Unknown parameters $\hat{A}_x$ are to be estimated. For ages $x > 0$ survival function $l(x+t)$ can be defined as a parabola within the elementary age interval $0 \le t \le 1$. 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

$$l(x+t) = l_x + (l_{x+1} - l_x)t + 6C_x(l_{x+1} - l_x)t(t-1), \qquad (5)$$

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

It is possible then to determine a polynomial of the fourth degree for the function of our interest $[l(x+t)]^2$ (see Appendix 1 for more details) and to derive the expression for $\hat{A}_x$ by using (4)

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

For a simple case where life table deaths are evenly distributed within an elementary age interval $A_x = \dfrac{1}{2}$ and $C_x = 0$ formula (6) yields

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

It implies that $\hat{A}_x = \dfrac{1}{2}$ if $q_x = 0$ and $\hat{A}_x = \dfrac{1}{3}$ 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 dying is higher, the decrease in $[l(x)]^2$ becomes considerably steeper than that of 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 (4) 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 (6) is also valid for an abridged life table if an elementary age interval $[x, x+n]$ parameter $A_x$ is defined as $\dfrac{({}_nL_x/n) - l_{x+n}}{l_x - l_{x+n}}$ and, therefore is varying between 0 and 1.

Formula (6) would not work in a proper way for $x = 0$ because during the first year of life $l(x)$ falls much steeper than it can be predicted by a quadratic polynomial. The use of the formula by

J.Borgois-Pichat (1951) instead of a parabola (Appendix 2) solves the problem for age 0 and results in

$$\hat{A}_0 = A_0\left(1 - q_0 \frac{3 + 0.831 \cdot A_0}{2 + q_0}\right). \qquad (7)$$

Let us, finally, find a solution for an open-age interval. For many countries mortality data running up to the highest ages are hardly available. For example, in the WHO Mortality Database (2001) the last age group is typically 85+. Fortunately, the Berkeley Mortality Database (Berkeley Mortality Database, 2000) provides 334 complete life tables for Japan, France, Sweden, and the USA with single-year age groups running up to 110. It gives the opportunity to establish a statistical relationship between "exact" values of $\hat{A}_{85}$ equal to $\frac{1}{(l_{85})^2}\sum_{x=85}^{109}\left[(l_{x+1})^2 + \hat{A}_x((l_x)^2 - (l_{x+1})^2)\right]$ and $e_{85}$. Linear regression returns the following relationships:

$$\hat{A}_{85} = -0.440 + 0.680 \cdot e_{85} \text{ for women, } \hat{A}_{85} = -0.227 + 0.626 \cdot e_{85} \text{ for men} \qquad (8)$$

Formulae (6), (7), and (8) give a set of parameters $\hat{A}_x$ for a correct numerical integration of the function $[l(x)]^2$.

Table 1 shows the magnitudes of errors depending on the type of input data (complete life tables, abridged life tables or abridged life tables with last age 85+) and the way of computation (with $A_x$ or $\hat{A}_x$). Complete life tables for Sweden for the years 1861, 1900, 1920, 1940, 1960, 1980, and 1995 were taken from the Berkeley Mortality Database (2001). Abridged life tables and abridged life tables with the last age group 85+ were made from complete life tables in a conventional way. Table 1 suggests that if the data of compete life tables without upper age limit are available, then it is not so important whether the original $A_x$ or modified $\hat{A}_x$ are used. Although in the former case where $G_0$ estimates are systematically lower than the correct ones, the deviation is very small. The difference between the functions $[l(x)]^2$ and $l(x)$ are greater at ages of rapid changes in mortality (infant and old ages). Consequently, the errors are somewhat greater for historical populations with a high proportion of deaths in infancy and for modern populations with a high proportion of deaths at very old ages.

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). Importantly, the error has tended to increase quite significantly in the last decades because the proportion of deaths occurring at ages above 85 is increasing steeply.

In all cases the use of modified parameters $\hat{A}_x$ reduce errors a great deal and in most cases they are very slight. In order to re-check our prior results on the data, which had not been used to estimate $\hat{A}_{85}$, we made another comparison. First, we computed $G_0$*100 values from 89 complete life tables (from the V.Kannisto's collection of life tables) 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 $G_0$*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 "exact" estimates was 0.189 for $A_x$-estimates and 0.014 for $\hat{A}_x$-estimates. For women the equivalent figures were 0.291 and 0.026, respectively.

Finally, Figure 3 shows the deviations of different $G_0 \cdot 100$ estimates from the "exact" estimates (computed from complete life tables with the last age 110 combined with using $\hat{A}_x$) for 334 life tables from the Berkeley Mortality Database. It also completely confirms the results of Table 1 on the basis of a large number of mortality schedules. In addition, Figure 3 shows that in most recent years estimates of $G_0$ computed from female abridged life tables with the last age group 85+ using $\hat{A}_x$ have been shifting slightly upwards from thir true values. This is a sign that the use of $\hat{A}_{85}$ can not replace data on mortality at ages above 85 if the proportion of life table deaths at ages over 85 is high. At present, respective error is small, but it would increase with time and in the future it will be necessary to use mortality data by age for ages above 85.

## 3. Decomposition of differences between Gini coefficients. Effects of age, cause of death and population composition.

### Age

When analysing changes in life expectancy in time or its variations across countries it is useful to be able to decompose observed differences by age and cause of death. This gives the opportunity of linking variations in overall life expectancy with differences in mortality rates in elementary age-cause categories to the overall difference in life expectancy.

For a similar reason the idea of decomposition of differences between two Gini coefficients arises. The age- and cause-specific components would show to what extent differences in elementary mortality rates at different ages and from different causes of death 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, USA, and France (Andreev, 1982, Arriaga, 1984, Pressat, 1985). The formula of decomposition by E.Andreev is exactly equivalent to that of 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 (Appendix 3). A continuous version of the method of decomposition by age was developed by J.Pollard (Pollard, 1982).

All these methods are based on the idea of standardisation or replacement (Kitagawa, 1955). 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. 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 mortality rate at age $x$ by respective mortality rate of the second population.

Let us, first, apply this general algorithm to life expectancy at birth and demonstrate that it leads to the conventional formula of decomposition. Let $\mu^{[x]}$ be the force of mortality function equal to the force of mortality of the second population $\mu'(t)$ if $t \le x$ 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

$$\delta_{0,x} = e_0(\mu^{[x]}) - e_0 = (L'_{0|x} - L_{0|x}) + (l'_x - l_x) \cdot e_x, \qquad (9)$$

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

$$e_0' - e_0 = \sum_{i=0}^{n-1} (\delta_{0,x_{i+1}} - \delta_{0,x_i}) = \sum_{i=0}^{n} \delta_i \qquad (10)$$

$\delta_i$ can be regarded as a contribution of the age interval $[x_i, x_{i+1}]$ to the overall difference between life expectancies at birth. Using (10) and (9) we easily come to the conventional formula of decomposition by Andreev and Pressat

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

On the other hand, it is clear that

$$\delta_i = e_0(\mathrm{M}^{[x_{i+1}]}) - e_0(\mathrm{M}^{[x_i]}), \qquad (11)$$

where $\mathrm{M}^{[x_i]}$ is a vector of age-specific mortality rates with elements $m_x'$ for $x \le x_i$ and $m_x$ for $x \ge x_i$. In fact, formula (11) can be considered as a general procedure for decomposition by age of a difference in aggregate measures based on the life table. It determines a stepwise replacement of one mortality schedule by another one, beginning from the youngest to the oldest age group.

Why does the replacement go from younger to older ages? Generally speaking, it can be organised differently. For example, it could go from old to young ages (Pollard, 1988) or in some other way. It seems, however, that a certain tradition has been already established and there is no reason to re-consider it at present. In addition, it is natural to move from young ages to old ages, following the ageing of the life table cohort.

It is possible now to obtain a decomposition formula for $G_0$. Using the definition (3) similarly to (9), the difference induced by mortality replacement at age $x$ and younger ages can be expressed as

$$\varepsilon_{0,x} = G_0(\mathrm{M}^{[x]}) - G_0 = \frac{\theta_0}{e_0} - \frac{\theta_{0|x}' + \theta_x (l_x')^2}{e_{0|x}' + e_x l_x'}, \qquad (12)$$

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 Gini coefficient by age group similar to (10) is

$$G_0' - G_0 = \sum_{i=0}^{n-1} (\varepsilon_{0,x_{i+1}} - \varepsilon_{0,x_i}) = \sum_{i=0}^{n} \varepsilon_i \qquad (13)$$

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

$$\varepsilon_i = G_0(\mathrm{M}^{[x_{i+1}]}) - G_0(\mathrm{M}^{[x_i]}) \qquad (14)$$

Formulae (12), (13) allow a difference to be split between two Gini coefficients according to age groups. Similar to life expectancy (Andreev, 1982, Pressat, 1985, see also Appendix 3), 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 (13), (14) twice and then to average the resulting age-specific components.

Formulae (12), (13) 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.

Procedure (14) can also be used for computation instead of (12), (13), but it is more laborious. However, we will demonstrate later that it 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 2 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 results for Gini coefficients are very consistent with those of Wilmoth and Horiuchi (1999). The total increase in life expectancy at birth is about 26 years for men and 30 years for women and the total decrease in $G_0 \cdot 100$ is about 24 for both sexes. About 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 dcrease in mortality at ages 65 and older. The overall decrease in Gini coefficient is distributed differently. The proportion of the decrease due to the youngest age group 0-14 is much higher (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 made a negative contribution of –5%.

--- Table 2 about here ---

Gini coefficient decreases when life table deaths concentrate around the average age at death. It appears to be very sensitive to mortality reductions in infants and children. On the other hand, the increase in life table deaths at very old ages makes the overall distribution of life table deaths more unequal as long as these ages are far apart from the average age at death. It does not mean that low mortality at old ages is bad for equality. It means that the relatively high remaining number of deaths at young and middle ages results in a relatively low average age at death. Using analogy with income one can say that it is bad that a relatively large number of people are still "poor" (deaths at young and middle ages) and it is not bad that some people are rich (deaths at old ages).

## Age and cause of death

With the help of the definition (3) formula (12) can be re-written without $e_{0|x}$ and $\theta_{0|x}$ as

$$\varepsilon_{0,x} = \frac{\theta_0}{e_0} - \frac{\theta_0' - \theta_x'(l_x')^2 + \theta_x(l_x')^2}{e_0 - l_x' e_x' + e_x l_x'} \tag{15}$$

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

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

Applying (13) and (15) to a small age interval $[x, x+\Delta x]$ after some transformations one can yield

$$\lambda_x = \frac{\varepsilon_{x,x+\Delta x}}{\Delta x} = \frac{(\mu_x - \mu'_x)l'_x}{[e'_0 + l'_x(e_x - e'_x)]^2}\{e_x[\theta'_0 + (l'_x)^2(\theta_x - \theta'_x)] - 2l'_x\theta_x[e'_0 + l'_x(e_x - e'_x)]\} = \quad (16)$$

$$= (\mu_x - \mu'_x)\eta_x.$$

Integrating (16) 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 $\mu_x = \sum_{j=1}^{m} \mu_{x,j}$ then

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

If in 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 (in relatively rare cases) a numerical integration would be necessary to compute cause-specific components according to

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

A comparison between the USA and the UK in male life expectancy at birth and Gini coefficient in the year 1997 is given as an example of decomposition by age- and cause of death (Figure 4). The life expectancy of men is very similar in both countries. The difference is only one year in favour of the UK (or 1.4%). However, there is a significant 16% difference in Gini coefficients in favour of the UK.

Figure 4 shows age- and cause-specific components of the difference. The advantage of the UK in male life expectancy (left panel of Figure 4) is mostly due to lower mortality rates regarding external causes of death (accidents and violence) at ages from 15 to 50 and, to some extend, to lower mortality rates regarding 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 regarding circulatory and respiratory diseases and cancers.

The weight of external causes of death at young adult ages is higher in the UK-USA difference in Gini coefficients (right panel of Figure 4) than that in the difference in life expectancies at birth. In addition, low mortality at old ages increases the level of Gini coefficient in the USA in comparison to the UK.

--- Figure 4 about here ---

Elimination of causes of death is another method for analysing 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).

Figure 5 shows temporal changes in the effects of elimination of leading classes of causes of death on $G_0$ for women in the UK in 1951-1996. Elimination of causes of death usually induces

a decrease in Gini coefficient. This is true for all classes of causes of death except cardiovascular diseases. Their elimination results in a substantial increase in $G_0$ because the distribution of life table deaths by age becomes more unequal. Since the mid-1970s respiratory diseases, which used to be associated with negative elimination effects in the 1950s-60s, joined the class of cardiovascular diseases. The reason for such a change is obvious. Respiratory diseases have transformed from an important cause of infant and child death into a cause of death of old people.

--- Figure 5 about here ---

In the 1950s the elimination of perinatal causes and congenital anomalies lead to the greatest decrease in $G_0$ among other causes of death. By the 1990s the effect of this major cause of infant deaths has been very much reduced. The same has happened to other common causes of mortality in childhood (infectious and respiratory diseases).

Nowadays, the greatest potential decrease in $G_0$ can be produced by elimination of external causes of death. The elimination effect for this class of causes has been relatively stable in time. The same is true for the elimination effect of cancer.

## Age and population group

Many types of mortality studies have to operate with mortality rates by age (or age and cause of death), but also by population group (social, ethnic, regional, etc.), each of which is characterised by its own vector of age-specific mortality rates. The decomposition of life expectancy and Gini coefficient by population group is an opportunity to link inter-individual and inter-group inequalities in the face of death.

Decomposition of aggregate measures based on cross-classified data has certain peculiarities in comparison to decomposition by variables (Das Gupta, 1994, Horiuchi et al., 2001). 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 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" can help to chose an appropriate 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 in aggregate measures 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 (11), (14).

Let $\mathrm{M} = \left\| m_{ij} \right\|$ be a matrix of mortality rates by age group $i$ and population group $j$ and $\mathrm{P} = \left\| p_{ij} \right\|$ be a matrix of the weights of groups in the overall population of age group $i$ ($\sum_j p_{ij} = 1, \forall i$). Apparently, 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 $\mathrm{M}^{[k]}$ consisting of elements $m'_{ij}$ for $i \le k$ and elements $m_{ij}$ for $i > k$. A corresponding matrix of population weights with replaced

rows (age groups) up to the age group $k$ $\mathrm{P}^{[k]}$ can be defined in a similar way. It is important to stress here that we always replace *entire rows* of matrices.

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

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

We consider two possibilities for a transition from $(\mathrm{M}^{[k-1]}, \mathrm{P}^{[k-1]}) \to (\mathrm{M}^{[k]}, \mathrm{P}^{[k]})$:

$(\mathrm{M}^{[k-1]}, \mathrm{P}^{[k-1]}) \to (\mathrm{M}^{[k]}, \mathrm{P}^{[k-1]}) \to (\mathrm{M}^{[k]}, \mathrm{P}^{[k]})$ or $(\mathrm{M}^{[k-1]}, \mathrm{P}^{[k-1]}) \to (\mathrm{M}^{[k-1]}, \mathrm{P}^{[k]}) \to (\mathrm{M}^{[k]}, \mathrm{P}^{[k]})$.
Accordingly, it is possible to get two versions of the components due to mortality rates (M-effect) and composition (P-effect) for the age group $k$:

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

and

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

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

$$\varepsilon_k^M = \varepsilon_k^{1,M} + \varepsilon_k^{2,M} = \frac{1}{2}\{[G_0(\mathrm{M}^{[k]}, \mathrm{P}^{[k]}) - G_0(\mathrm{M}^{[k-1]}, \mathrm{P}^{[k]})] + [G_0(\mathrm{M}^{[k]}, \mathrm{P}^{[k-1]}) - G_0(\mathrm{M}^{[k-1]}, \mathrm{P}^{[k-1]})]\} \quad (17)$$

$$\varepsilon_k^P = \varepsilon_k^{1,P} + \varepsilon_k^{2,P} = \frac{1}{2}\{[G_0(\mathrm{M}^{[k-1]}, \mathrm{P}^{[k]}) - G_0(\mathrm{M}^{[k-1]}, \mathrm{P}^{[k-1]})] + [G_0(\mathrm{M}^{[k]}, \mathrm{P}^{[k]}) - G_0(\mathrm{M}^{[k]}, \mathrm{P}^{[k-1]})]\} \quad (18)$$

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 $(\mathrm{M}^{[k-1]}, \mathrm{P}^{[k-1]}) \to (\mathrm{M}^{[k]}, \mathrm{P}^{[k-1]})$ and $(\mathrm{M}^{[k-1]}, \mathrm{P}^{[k]}) \to (\mathrm{M}^{[k]}, \mathrm{P}^{[k]})$. In our prior consideration it was very simple: row $k$ was to be replaced entirely. However, to obtain the effect of mortality rate in the particular population group $j$ and 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 different 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 are making M-replacement for the age group 20 and there are two population groups 1 and 2 then the effect of mortality rate in population group 1 would be

$$\varepsilon_{20}^{M(P1)} = \frac{1}{2}\{\frac{1}{2}[G_0(\|m'_{20,1}, m_{20,2}\|, \mathrm{P}^{[20]}) + G_0(\|m'_{20,1}, m'_{20,2}\|, \mathrm{P}^{[20]})] - G_0(\mathrm{M}^{[19]}, \mathrm{P}^{[20]})\} +$$
$$+ \frac{1}{2}\{\frac{1}{2}[G_0(\|m'_{20,1}, m_{20,2}\|, \mathrm{P}^{[19]}) + G_0(\|m'_{20,1}, m'_{20,2}\|, \mathrm{P}^{[19]})] - G_0(\mathrm{M}^{[19]}, \mathrm{P}^{[19]})\}$$

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

$$\varepsilon_{20}^{M(P2)} = \frac{1}{2}\{[G_0(\|m_{20,1}, m'_{20,2}\|, \mathrm{P}^{[20]}) + G_0(\|m'_{20,1}, m'_{20,2}\|, \mathrm{P}^{[20]}) - G_0(\mathrm{M}^{[19]}, \mathrm{P}^{[20]})\} +$$
$$+ \frac{1}{2}\{[G_0(\|m_{20,1}, m'_{20,2}\|, \mathrm{P}^{[19]}) + G_0(\|m'_{20,1}, m'_{20,2}\|, \mathrm{P}^{[19]}) - G_0(\mathrm{M}^{[19]}, \mathrm{P}^{[19]})\}$$

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

Table 3 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 levels. Three processes contributed to this favourable change: the work of education itself, the natural replacement of older generations with low average educational levels by younger people, and, to some extent, migration into Russia of people with relatively high levels of education from other parts of the USSR.

Table 4 suggests that between 1979 and 1989 life expectancy within the age interval 20 to 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 and a decrease in Gini coefficient were very modest. Interestingly, for the whole population the increase in life expectancy (1.4 years) and decrease in Gini coefficient (-2.4) were substantially greater than for those in 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).

A widening of inter-group differences in the life expectancy of Russian men in 1979-89 coincides with a substantial decline in Gini coefficient (Table 4). This illustrates a difference in the meanings of inter-individual inequality in length of life and length-of-life differentials in length of life 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 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. So, the educational gradient is much more pronounced in terms of the degree of inequality the distribution of individuals by length of life.

Table 5 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 and effects of educational composition. Mortality effects are also given for each educational group. 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 and Gini coefficient. 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 a low education was very small (Table 4), its contribution to overall improvement is greater than that for university education because low education has a higher weight in the population.

Finally, it is important to note that the algorithm for the decomposition of differences given in this section can be applied not only to life expectancy or Gini coefficient, but to any aggregate measure based on the life table.

## 4. Variations 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, Wilmoth and Horiuchi, 1999, Llorka et al., 1998). First, during the 20th century the inequality (or variability) in length of life was declining, mirroring the increase in average length of life. Second, during the last decade this correlation has become weaker since life expectancy has continued increasing, while the decline in inequality 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 independent from a choice of inequality indices.

An "evolutionary" explanation of this regularity was given by Wilmoth and Horiuchi (1999). The historical lowering of mortality rates was much more pronounced in the young than in old ages. Therefore, life table deaths have been concentrated around the average length of life. After a certain point (in the 1950s, 60s or 70s depending on the country) at which mortality at young ages had already been reduced to low values, further reduction was unable to significantly effect a general degree of inequality. In addition, in the 1980s-90s the mortality decline in countries with low mortality was more pronounced at old ages than at young or middle ages. This process (as it was shown earlier) produces positive contributions to inequality as measured by the Gini coefficient.

It 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 many 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 given here) with the values of Gini coefficient and life expectancy for about 45 countries for the period of 1960s-1990s suggest that if one considers a wide variety of populations with very different life expectancies then the association between life expectancy at birth and Gini coefficient for the full range of ages would be very tight, with coefficients of correlation 0.88-0.95 depending on the selection of countries and years. If only countries with comparable levels of mortality are selected then this correlation is smaller due to the country's peculiarities in respect to public health and social situations.

Figure 5 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. Correlation coefficient between life expectancy and Gini coefficient by country is 0.7 for men and 0.6 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 male life expectancy is 73.6 with Gini coefficient 12.3, while in Ireland the equivalent figures are 73 and 10.6. In Chile, the USA, Cuba and Singapore the values of Gini coefficient 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 their movement 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 progression in the lengthening of life and a reduction of inequality in length of life.

---- Figure 6 about here ----

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. Inequality in length of life in Japan has always been lower than that in the USA after the level of 65 years of male life expectancy. In Spain inequality in length of life 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 in Japan, Russia or Spain. Male life expectancy in these countries has been rather similar with only one year in favour of the UK during almost the whole period of observation. There is, however, a considerable difference between the two countries in Gini coefficient, which appears to be unusually low in the UK. In the UK the level of inequality in length of life is always somewhat lower than in other countries with similar levels of average length of life.

---- Figure 7 about here ----

Finally, one can see periods of flattening in Gini coefficient in the USA and Spain in the 1980s and early 1990s (Figure 6). Figure 7 presents a zoomed image of changes in male and female life expectancies and Gini coefficients in the USA in 1980-1996. For men life expectancy at birth continues increasing over the period, while 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 more clear for men because life expectancy at age 15 increases, while Gini coefficient for ages above 15 increases too, 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 Gini coefficient declines.

Decomposition of the increase in $G_{15}$ for US men between 1983 (beginning of the increase) and 1989 (a point of maximum) shows that it can be explained by two major factors. First, deterioration in mortality of adults at ages from 15 to 44 due to AIDS. 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 overall shape of distribution of deaths by age.

---- Table 6 about here ----

In Spain the nature of the increase in Gini coefficient since 1985 was similar to that in the USA. Llorka et al. (1998) explains the increase in Gini coefficient after 1985 mostly by mortality from AIDS. Our decomposition for ages 15+ confirms this finding.

## *Acknowledgements*

Preparatory work related to the present study was done within the framework of the Global Health Equity Initiative of the Rockefeller Foundation.

We would like to express our sincere gratitude to Professor Sudhir Anand, who motivated our interest in the subject and with whom some of its ideas had been discussed during 1997-99. We are also grateful to John Wilmoth and his team at the University of California in Berkeley for

their great work on the Berkeley Mortality Database, which was extensively used in our study as a rich source of information on variations in mortality patterns.

Finally, we are very 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.

## Appendix 1

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

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

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

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

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 2

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

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

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

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

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

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

$$\alpha = q_0 - \beta \cdot C,$$
$$\beta = \frac{A_0 q_0}{C - S_3} \qquad \text{(C)}$$

$[l(t)]^2$ 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

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

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

From the definition (4) it follows that $\hat{A}_0 = \dfrac{\int_0^1 l(x)dx - (l_1)^2}{1-(l_1)^2}$. Using this expression, (D), values of the constants ln(366)=5.9026, C=205.6541, $S_3$=130.9208, $S_6$=20106.74 and equations (C) for parameters $\alpha$ и $\beta$ 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 3*

According to Andreev (1982):

$$\begin{aligned} e_0^2 - e_0^1 &= \sum_x {}_n\delta_x^{2,1} = \sum_x \left[l_x^2(e_x^2 - e_x^1) - l_{x+n}^2(e_{x+n}^2 - e_{x+n}^1)\right], \\ e_0^1 - e_0^2 &= \sum_x {}_n\delta_x^{1,2} = \sum_x \left[l_x^1(e_x^1 - e_x^2) - l_{x+n}^1(e_{x+n}^1 - e_{x+n}^2)\right], \end{aligned} \qquad \text{(A)}$$

Generally speaking, ${}_n\delta_x^{2,1} \neq -{}_n\delta_x^{1,2}$. That is to say, the age-components of a difference between two life expectancies depend on the permutation of populations.

To avoid this problem Andreev (1982) and Pressat (1985) suggested using

$${}_n\varepsilon_x = \frac{1}{2}({}_n\delta_x^{2,1} - {}_n\delta_x^{1,2}) \qquad \text{(B)}$$

for a "symmetrical" decomposition of $e_0^2 - e_0^1$. It is easy to see that (A) and (B) are equivalent to expression (2) from section 1.

The formula by E.Arriaga (1984) replicated in the recent textbook by Preston *et al*. (2001) is written in the following form:

$$e_0^2 - e_0^1 = \sum_x {}_n\Delta_x^{2,1} = \sum_x \left[ l_x^1\left(\frac{{}_nL_x^2}{l_x^2} - \frac{{}_nL_x^1}{l_x^1}\right) - T_{x+n}^1\left(\frac{l_x^1}{l_x^2} - \frac{l_{x+n}^1}{l_{x+n}^2}\right)\right] \qquad \text{(C)}$$

Taking into account $T_x = l_x \cdot e_x$ and ${}_nL_x = l_x e_x - l_{x+n}e_{x+n}$ one can express ${}_n\Delta_x^{2,1}$ in a simpler form:

$${}_n\Delta_x^{2,1} = l_x^1(e_x^2 - e_x^1) - l_{x+n}^1(e_{x+n}^2 - e_{x+n}^1) \qquad \text{(D)}$$

Comparison of (A) and (D) shows that ${}_n\Delta_x^{2,1} = -{}_n\delta_x^{1,2}$. The latter means that the difference between the components by Arriaga and those by Andreev-Pressat does not exist if one uses symmetrical components (B).

## *References*


Allison P.D. (1978) Measures of inequality. *American Sociological Review* 43: 865-880.

Anand S., Diderichsen F., Evans T., Shkolnikov V., 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, Abbas Bhuiya. New York: Oxford University Press, pp. 49-67

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*. Mimeo, Harvard Center for Population and Development Studies.

Anand, S., Diderichsen, F., Evans, T., Shkolnikov, V., 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, Abbas Bhuiya. New York: Oxford University Press, pp. 49-67 (in press)

Andreev, E.M. (1982) Metod komponent v analize prodoljitelnosty zjizni. [The method of components in the analysis of length of life]. *Vestnik Statistiki*, No 9, p. 42-47.

Arriaga, E. (1984) Measuring and explaining the change in life expectancies. *Demography*, vol. 21, No 1, p. 83-96.

Atkinson, A.B. (ed) 1980. *Wealth, Income and Inequality*, 2nd edn, Oxford: Oxford University Press.

Atkinson, A.B. 1970. On the measurement of inequality. *Journal of Economic Theory* 2: 244-63.

Black, D., Morris, J.N., Smith, C., Townsend ,P. 1980. Inequalities in health (The Black Report). In: Townsend, P., Whitehead, M., Davidson, N. (eds) *Inequalities in Health: The Black report and The Health Divide*. 2nd Edition London. Penguin. 1992.

Bourgeois-Pichat, J. 1951. La mesure de la mortalite infantile. Principes et methodes. *Population*, 2: 233-248

*Demographic Surveillance System – Matlab. 1998*. Volume 28. Scientific Report No 82, January. International Centre for Diarrhoeal Disease Research, Bangladesh.

Fries, J.F. 1980. Aging, natural death and the compression of morbidity. New England Journal of Medicine

Hanada K. 1983. A formula of Gini's concentration ratio and its application to life tables. *Journal of Japan Statistical Society*, 13: 95-98.

Horiuchi H., Wilmoth J.R., Pletcher S.D. 2001. A general method of decomposition analysis. *The manuscript*.

Illsey, R., Le Grand, J. 1987. The measurement of inequality in health. In: *Health and Economics*. Ed by A.Williams. London: Macmillan: 13-36.

Kannisto V., Lauritsen, A., Thatcher, R., 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.

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.

Le Grand, J. and Rabin, M. 1986. Trends in British health inequality: 1931-83. In: Culyer, A.J. and Jonsson, B., *Public and private health services*. Oxford: Blackwell.

Le Grand, J., Rabin, M. 1986. Trends in British health inequality: 1931-83. In: Culyer, A.J. and Jonsson, B., *Public and private health services*. Oxford: Blackwell.

Leclerc, A., Lert, F., 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.

Llorka, 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., Brown, S. 2001. Reconsidering mortality compression and deceleration: an alternative model of mortality rates. *Demography*, vol. 38, 1, pp. 79-95.

Matlab Report. 1996. *Demographic Surveillance System – Matlab. Registration of Vital Events.* Volume 27, Scientific Report No 76, December.

Myers, G.C., Manton, K.G. 1984. Compression of mortality: myth or reality? *The Gerontologist*, 24(4): 346-353.

Pollard J.H. 1988. On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography* 25: 265-276.

Pollard, J.H. (1982) The expectation of life and its relationship to mortality. *Journal of the Institute of Actuaries*, vol. 109, Part II, No 442, p. 225-240.

Pressat, R. (1985) Contribution des écarts de mortalité par âge à la différance des vies moyennes. *Population*, No 4-5, p. 766-770.

Preston, S.H., Heuveline, P., Guillot, M. (2001). *Demography. Measuring and Modelling Population Process.* Blackwell Publishers. Oxford, Malden, Massachusetts.

Sen, A. 1973. *On economic inequality*. Oxford: Clarendon Press.

Shkolnikov (Vladimir), Meslé (France) et Vallin (Jacques) 1996. Health crisis in Russia. *Population: An English Selection*, 8, pp. 123-190.

Shkolnikov V., Leon D., 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,* Vol. 47, No 3, pp. 357-369.

The Berkeley Mortality Database. 2001. http://www.demog.berkeley.edu/wilmoth/mortality. Last visited 25.05.2001.

WHO Mortality Database. 2001. http://www.who.int/whosis/mort/download.htm Last visited 25.05.2001.

Wilmoth, J.R., Horiuchi, S. 1999. Rectangularization revisited: variability in age at death within human populations. *Demography*, vol. 36, 4, pp. 475-495.

Wilmoth, J.R., 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.

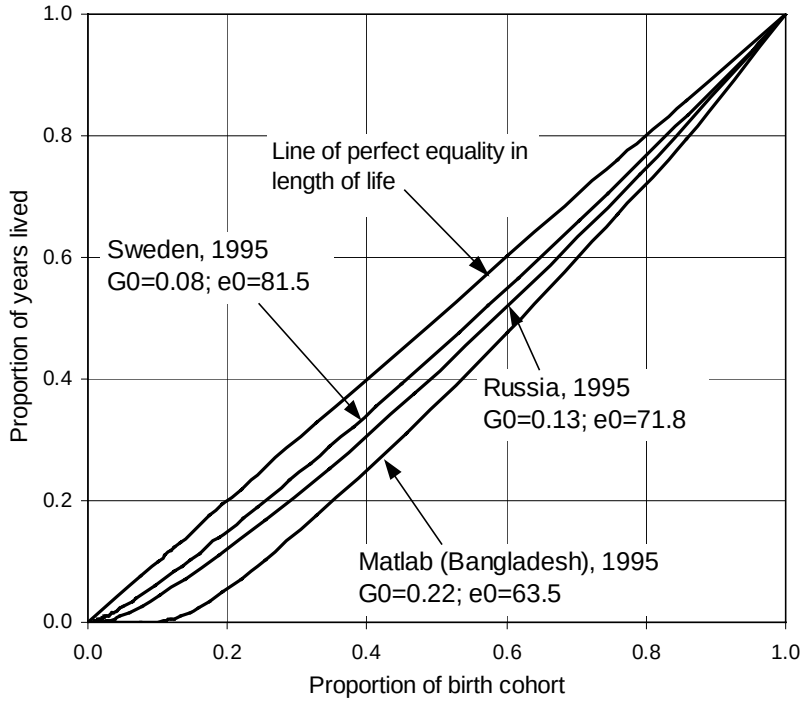



**Figure 1. Lorenz curves for three female populations with different levels and age distributions of mortality.**

Sources: Data for computations for Sweden are extracted from the Berkeley Mortality Database. Our own estimates are based on the original Goskomstat data on deaths and population by age for Russia. Matlab Report, 1996 for Bangladesh.

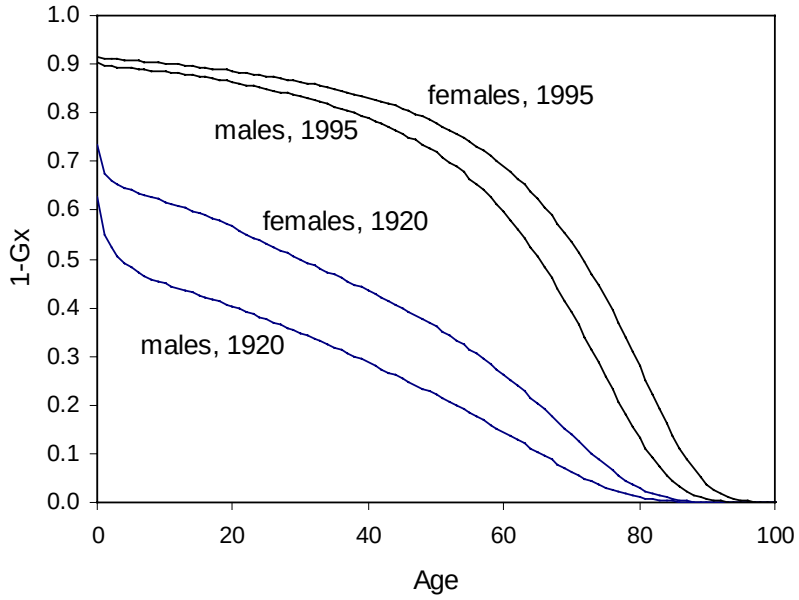
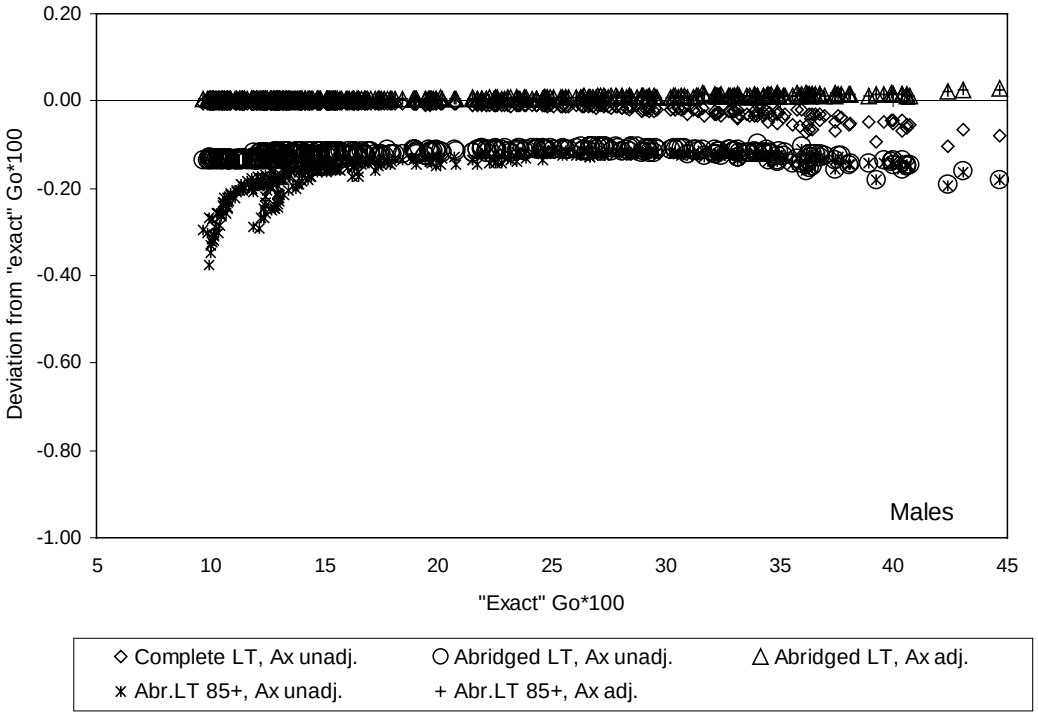



**Figure 2. Age curves *1-$G_x$* for Swedish male and female life tables for 1920 and 1995.**
Sources: Data for computations are extracted from the Berkeley Mortality Database (2001).

**Table 1. Life expectancy at birth and different estimates of the 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 $G_0$*100 from: | | | | | | Errors in $G_0$*100 estimates: | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Complete LT with $\hat{A}_x$ "exact" | Complete LT with $A_x$ | Abridged LT with $A_x$ | Abridged LT with $\hat{A}_x$ | Abridged LT, 85+ with $A_x$ | Abridged LT, 85+ with $\hat{A}_x$ | Complete LT with $A_x$ | Abridged LT with $A_x$ | Abridged LT with $\hat{A}_x$ | Abridged LT, 85+ with $A_x$ | Abridged LT, 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).



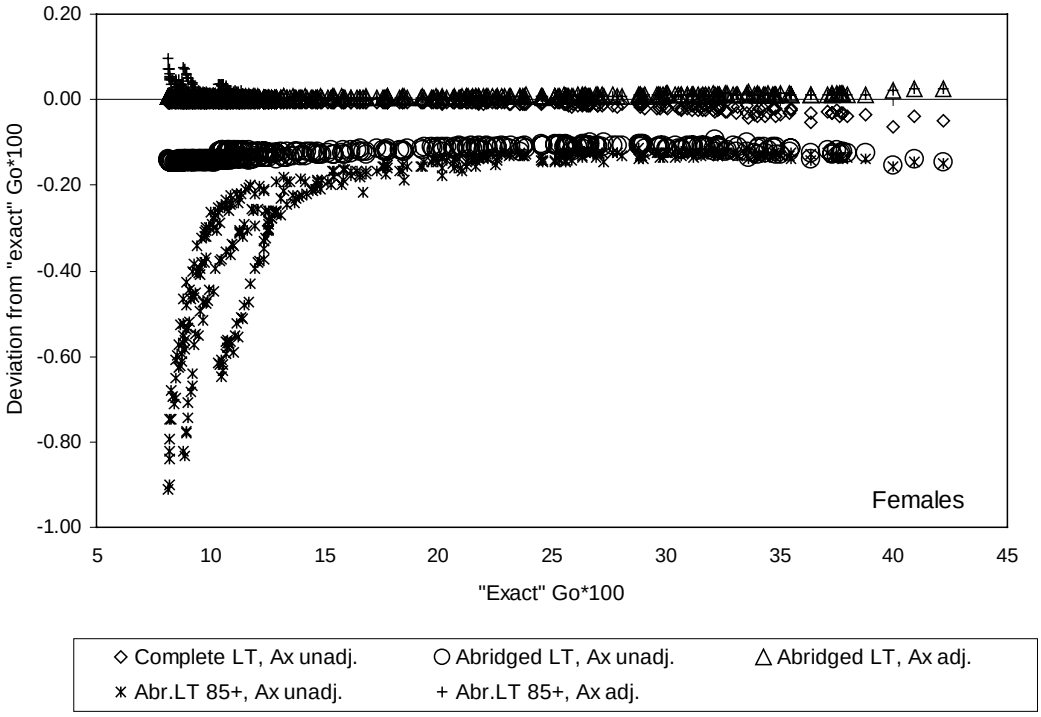




**Figure 3. Differences between "exact" $G_0$ and its estimates for 334 life tables computed from complete life tables, abridged life tables, abridged life tables with the last age 85+ before and after modification of the life table $A_x$. ("Exact" $G_0$ is computed from complete life tables with the last age 110 using $\hat{A}_x$).**

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

**Table 2. Age-specific contributions to the increase in life expectancy at birth and 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 |

** e0(1900)=46.4, e0(1995)=72.73*
*G0(1900)=36.73, G0(1995)=12.71*
Sources: Data for computations are extracted from the Berkeley Mortality Database (2001)

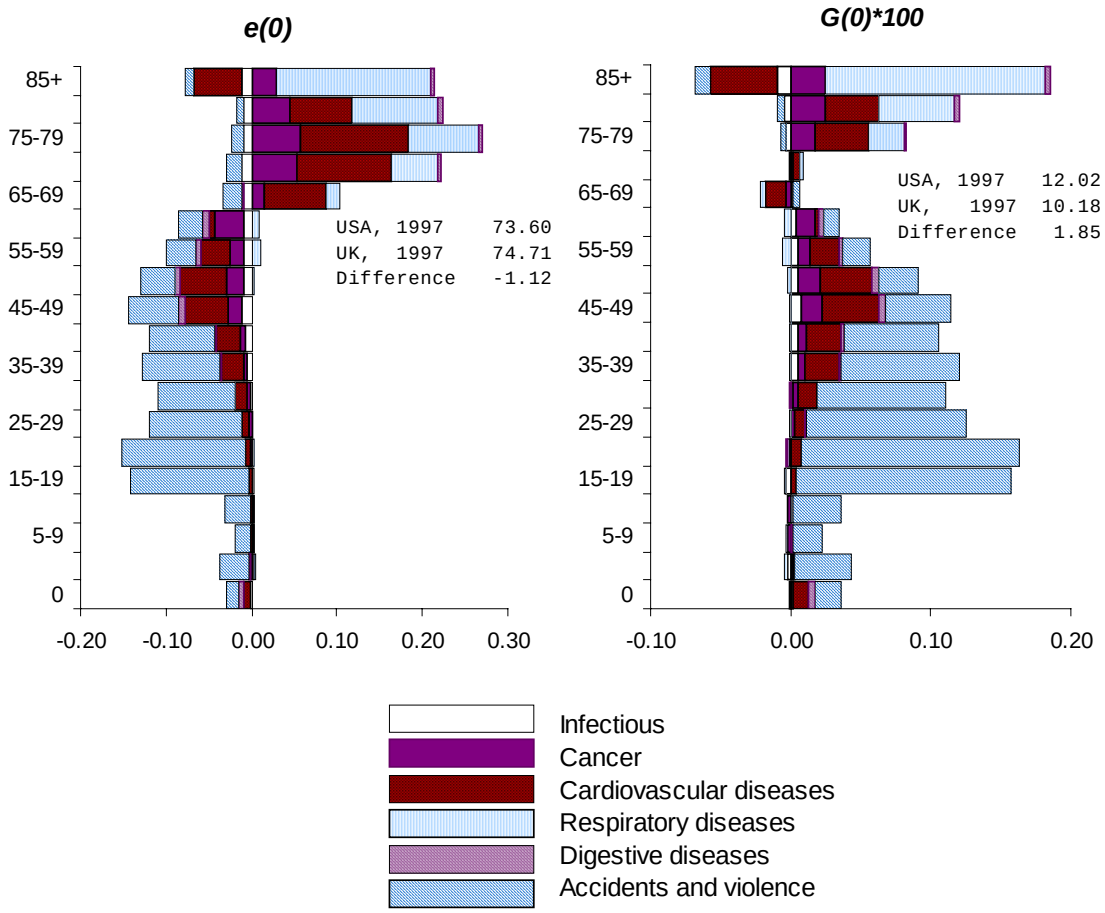



**Figure 4. Decompositions of the differences in life expectancy at birth and in Gini coefficient between the UK and the USA by age and cause of death: male populations, 1997.**

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

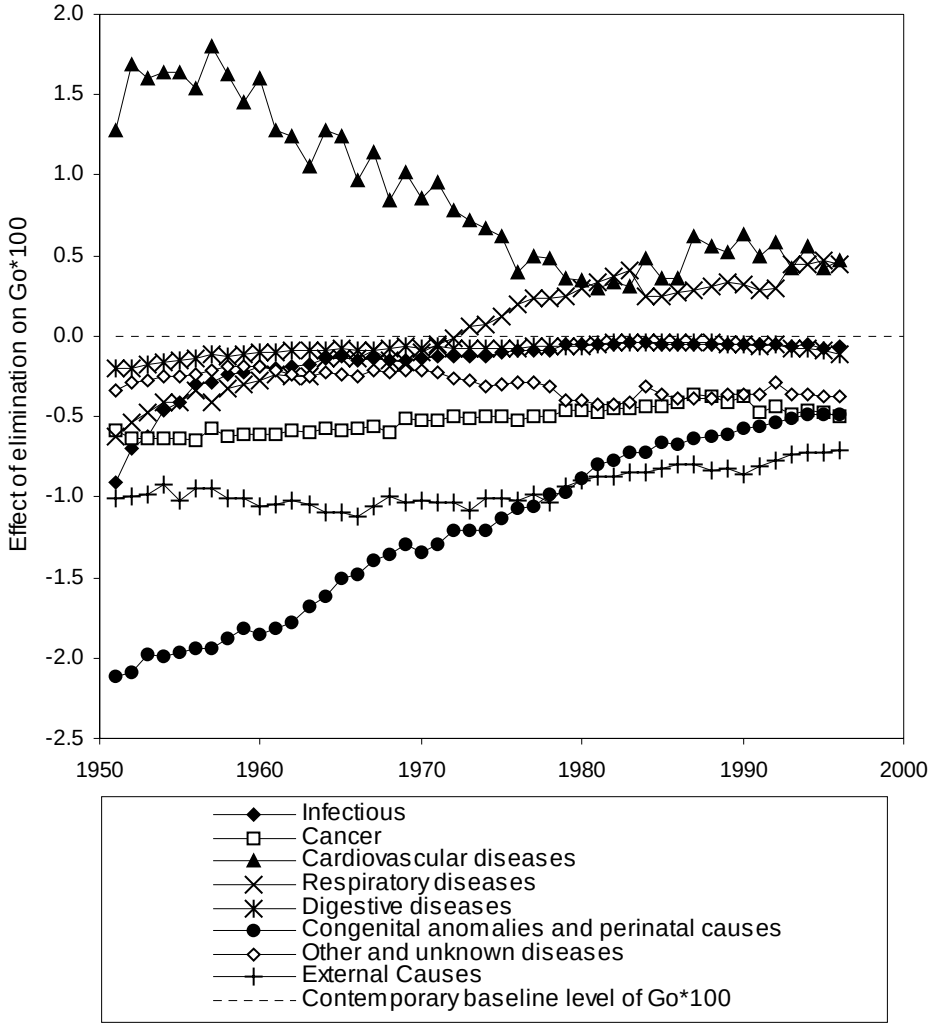



**Figure 5. Effects of elimination of causes of death on Go for women in the UK in 1951-1996**

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

**Table 3. Educational composition of the Russian male population by age in 1979 and 1989.**

| Age group | University education | Secondary education | Low 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 Goskomstat's tables of deaths and population by age and educational status.

**Table 4. 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(20-64) | | | G(20-64) | | |
|---|---|---|---|---|---|---|
| | 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 |
| Low education | 36.70 | 37.04 | 0.34 | 16.04 | 15.57 | -0.47 |

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

**Table 5. 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) | | | | Component due to changing educational composition (P-effects) | Total |
|---|---|---|---|---|---|---|
| | University education | Secondary education | Low education | Total | | |
| | e(20-64) | | | | | |
| 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 |
| | G(20-64) | | | | | |
| 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.

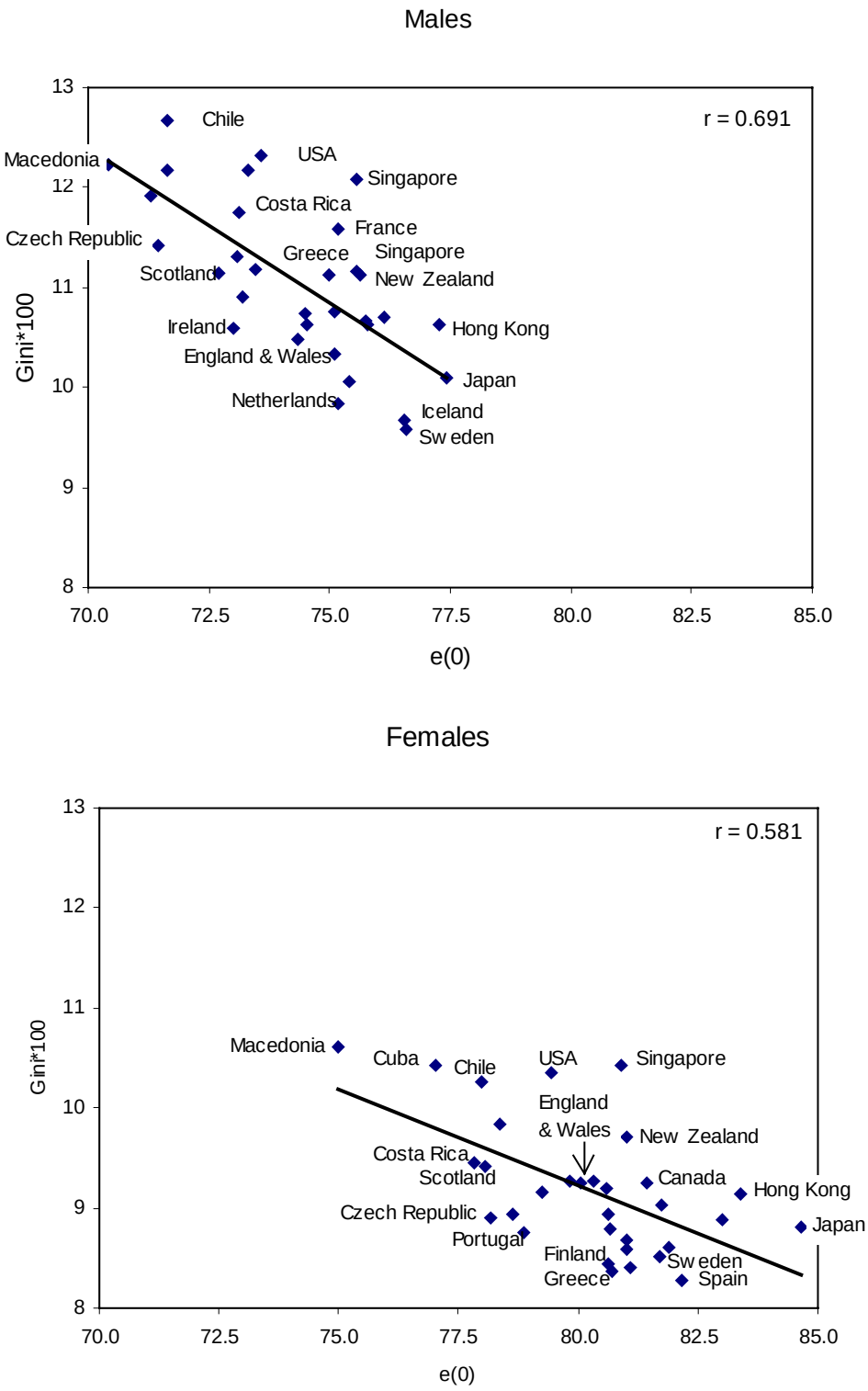



**Figure 5. Relationship between life expectancy and Gini coefficient in 1996-99 for 31 countries with male life expectancy higher than 70 years.**

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

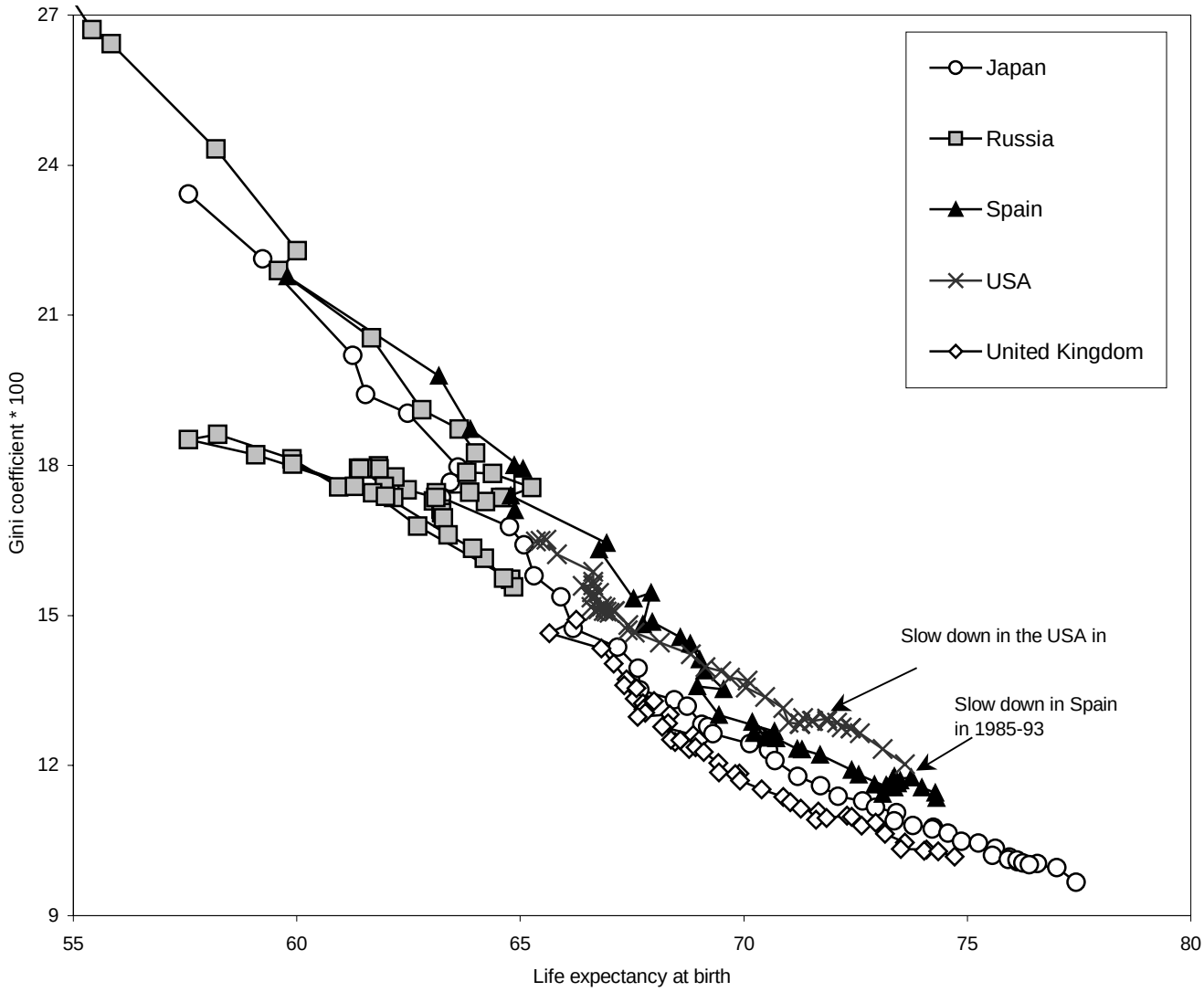



**Figure 6. Trajectories of five countries in coordinates $e_0$ and $G_0$ for the male populations of Japan, Russia, Spain, USA, and the UK in 1950-99.**

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

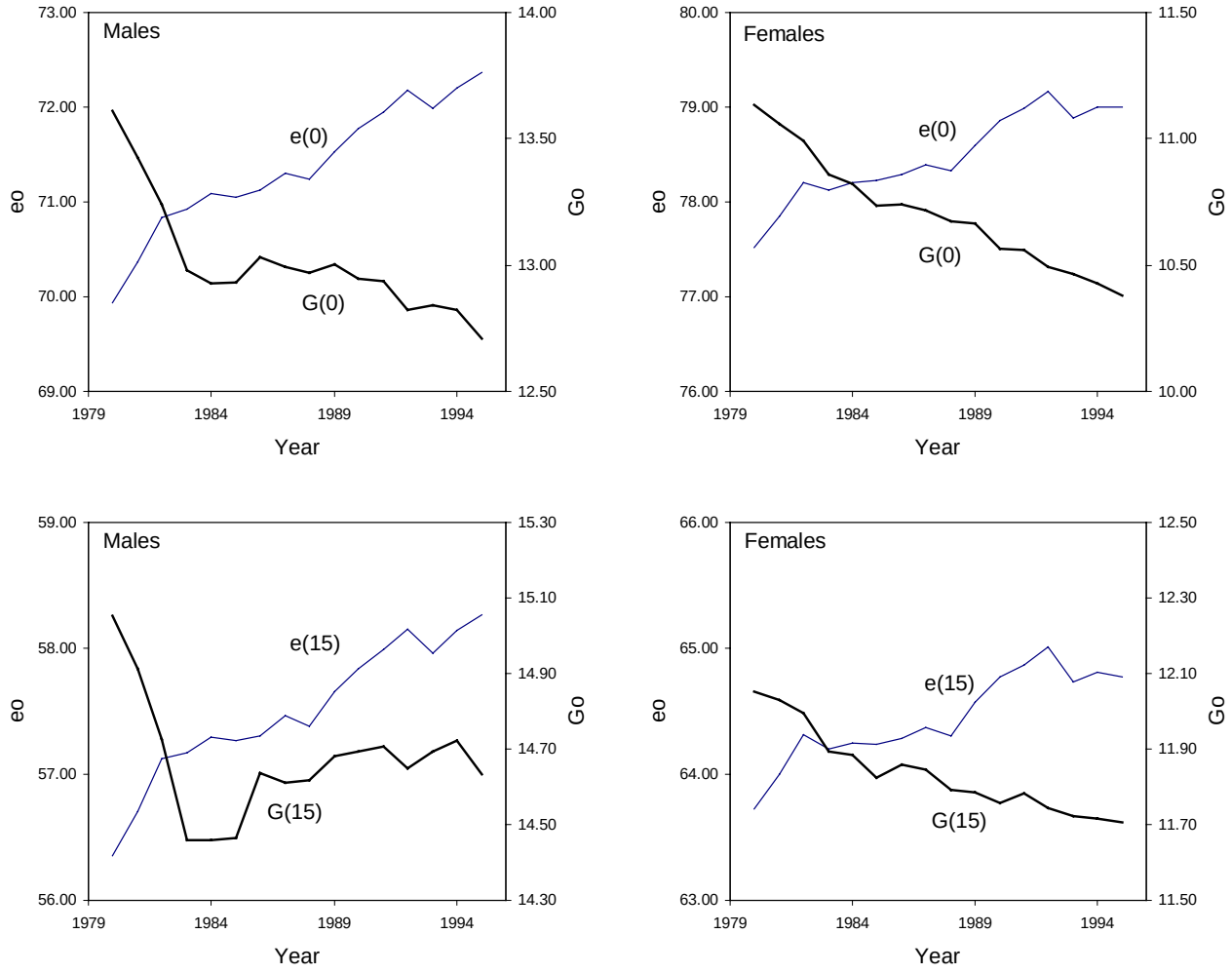



**Figure 7. Trends in life expectancy and Gini coefficient (age 0 and 15) for men and women in the USA in 1980-95.**

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

**Table 6. Components of the increase in $G_{15}$ *100 for the US men in 1983-1991* 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 |

* Gini coefficient*100 for ages 15+ was 14.6 in 1983 and 14.9 in 1999

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