

Max-Planck-Institut für demografische Forschung Max Planck Institute for Demographic Research Konrad-Zuse-Strasse 1 · 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 2014-010 NOVEMBER 2014

Assessment of cross-sectional and longitudinal components of a difference with an algorithm of contour replacement

Dmitri Jdanov (jdanov@demogr.mpg.de) Vladimir M. Shkolnikov (shkolnikov@demogr.mpg.de)

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

Assessment of cross-sectional and longitudinal components of a difference with an algorithm of contour replacement

Dmitri Jdanov and Vladimir M. Shkolnikov

Abstract

This study proposes a new decomposition method which permits a difference in an aggregate measure at a final time point to be split into additive components corresponding to the initial differences in the event-rates of the measure and differences in trends in these underlying event-rates. The method is an extension of the existing algorithm of stepwise replacement. We provide a full description of the method and two examples of its application with mortality and fertility data. The outcome of these two decompositions are: 1) age-specific contributions of initial conditions and trends to the Japan-USA life expectancy gap in 2009 with a reference time point at 1970; and 2) age-specific contributions of initial conditions and trends to the difference in Czech-Russian age variations in fertility for the 1970 female birth cohorts with reference to the 1955 cohorts.

Keywords

Decomposition, demographic change, stepwise replacement, mortality, fertility, aggregate demographic measure.

Background

Decomposition analysis provides insight into the reasons for a difference in an aggregate demographic index. This difference can be either the difference in the index between two populations at a given time, or a temporal change in the index for one population measured over two time periods. A number of techniques have been developed to split these differences into additive components produced by differences (or changes) in the elementary event-rates according to age or other dimensions such as cause of death, birth order, sub-population, etc. (Andreev, Shkolnikov, Begun, 2002; Horiuchi, Wilmoth, Pletcher, 2008; Beltran-Sanchez, Preston, Canudas-Romo, 2008; Nau and Firebaugh 2012; Caswell 2001; van Raalte and Caswell 2013; Andreev 1982; Arriaga 1984; Pressat 1985). In this way the relative importance of different ages, causes of death, birth orders or other factors to changes or differences in the aggregate index can be directly determined. In many cases there are good reasons to present decomposition results in a brief manner by summing the numerous elementary components into broader categories.

The present study proposes a general decomposition method that can be used to split an inter-population difference not only by a demographic dimension (e.g. age¹), but also according to the impact of 1) initial conditions and 2) time trends. The proposed discrete method for decomposition is an extension of the earlier general algorithm of stepwise replacement (Andreev, Shkolnikov, Begun, 2002).

Imagine that an aggregate measure (say, life expectancy) of two populations is measured at two time points (Figure 1). In order to understand the reasons for the inter-population difference at time *T*, one might want to assess the contributions of different ages to the total between-population difference at this later time point. This is a conventional decomposition

¹ We refer to age as the most common and natural dimension. But it can be any state (birth order, cause of death or other) that makes up the vector of rates.

that can be done with any number of the above-mentioned techniques. Yet it is also clear that every age component depends on initial differences between respective event rates (initial conditions) and on the temporal changes in these rates in the two populations (trends) going back to some earlier time period, *t*.

At first glance, it would seem natural to disentangle the initial conditions and trend components by summing: 1) an age decomposition of the between-population difference Δ_{ab} at the first time point; and 2) an age decomposition of the temporal changes δ_{Aa} , δ_{Bb} between the two time points within each of the two populations. It appears, however, that the sum of the initial between-population *x*-component and the difference between the two trend *x*-components is not equal to the cross-sectional *x*-component of the total between-population difference at the final time point.

Let us define the decomposition problem in a more formal way. Assume that the demographic measure of interest for a population *A* is a function of a vector of age-specific event rates:

$$E = f(\mathbf{m}_A)$$
,
 $\mathbf{m}_A = [m_A(x_1), \dots, m_A(x_i), \dots, m_A(x_n)]^2$

For the two populations A and B at time T measure E has values $f(\mathbf{m}_A)$ and $f(\mathbf{m}_B)$, respectively (Figure 1). The final between-population difference is

$$\Delta_{AB} = f(\boldsymbol{m}_A) - f(\boldsymbol{m}_B).$$

This final difference is being considered as resulting from initial differences between agespecific event rates in the two populations m_a and m_b and temporal changes in the agespecific rates (from m_a to m_A and from m_b to m_B).

Thus, the *decomposition task* is to split the final difference Δ_{AB} into age-specific contributions due to initial differences in age-specific mortality and age-specific contributions due to different (within country) age-specific mortality trends.

In the next part we show how the general algorithm of stepwise replacement can be extended for completing this decomposition task.

² For simplicity, the elementary age interval $[x_i, x_{i+1}]$ is designated by its beginning age x_i in this and later formulae.



Figure 1. Cross-sectional differences and longitudinal changes in an aggregate demographic measure.

Method

The vector of age-specific rates in a population *B* after replacement of *i* first elements of this vector by corresponding age-specific rates from vector *A* is

$$\boldsymbol{m}_{AB}^{[i]} = [m_A(x_1), \dots, m_A(x_i), m_B(x_{i+1}), \dots, m_B(x_n)]$$
 with $\boldsymbol{m}_{AB}^{[0]} = \boldsymbol{m}_B$

According to Andreev et al. (2002), the component of the total difference Δ_{AB} produced by the elementary difference between A and B in event rates at age [x_i , x_{i+1}) is

$$\Delta_{AB}^{i} = f\left(\boldsymbol{m}_{AB}^{[i]}\right) - f\left(\boldsymbol{m}_{AB}^{[i-1]}\right), \ i = 1, \dots, n.$$

$$\tag{1}$$

The basic equation of the stepwise replacement algorithm for a conventional decomposition of a between population difference Δ_{AB} is

$$f(\boldsymbol{m}_{A}) - f(\boldsymbol{m}_{B}) = \sum_{i=1}^{n} \left[f\left(\boldsymbol{m}_{AB}^{[i]}\right) - f\left(\boldsymbol{m}_{AB}^{[i-1]}\right) \right] = \sum_{i=1}^{n} \Delta_{AB}^{i}$$
(2)

Equation (2) reflects the process of replacement of elements of vector \mathbf{m}_{B} by elements of vector \mathbf{m}_{A} . Because there is no preference for the direction of the replacement, the final age-specific components should be calculated as an average

$$\Delta_{AB} = \sum_{i=1}^{n} \frac{1}{2} \left(\Delta_{AB}^{i} - \Delta_{BA}^{i} \right) \tag{3}$$

In earlier studies (Andreev, Shkolnikov, Begun, 2002; Shkolnikov, Valkonen, Andreev, Begun, 2001) we showed that in the case of life expectancy the stepwise replacement decomposition (1)-(3) is equivalent to the age decompositions by well-known formulae, which were independently derived in the 1980s by Andreev (1982), Arriaga (1984), and Pressat (1985).

Equations similar to (1), (2), and (3) for the trend decomposition of change δ_{Aa} from *a* to *A* (Figure 1) are

$$\delta_{Aa}^{i} = f\left(\boldsymbol{m}_{Aa}^{[i]}\right) - f\left(\boldsymbol{m}_{Aa}^{[i-1]}\right), \ i = 1, \dots, n \quad ,$$

$$(1a)$$

$$\delta_{Aa} = f(\boldsymbol{m}_{A}) - f(\boldsymbol{m}_{a}) = \sum_{i=1}^{n} \left[f\left(\boldsymbol{m}_{Aa}^{[i]}\right) - f\left(\boldsymbol{m}_{Aa}^{[i-1]}\right) \right] = \sum_{i=1}^{n} \delta_{Aa}^{i}$$
(2a)

$$\delta_{Aa} = \sum_{i=1}^{n} \frac{1}{2} \left(\delta_{Aa}^{i} - \delta_{aA}^{i} \right) \tag{3a}$$

As mentioned in the previous section, in general, the final between-population elementary *x*-components Δ_{AB}^{i} cannot be expressed via the conventional *x*-components:

$$\Delta_{AB}^{i} \neq \Delta_{ab}^{i} + \left(\delta_{Aa}^{i} - \delta_{Bb}^{i}\right), \quad i = 1, 2, \dots, n.$$

This is related to the path-dependence of decomposition outcomes (for more details see Andreev, Shkolnikov, Begun, 2002 and Horiuchi, Wilmoth, Pletcher, 2008).

Our present decomposition task implies a decomposition of the difference between states *A* and *B* conditioned on the past difference between *a* and *b* and the temporal changes from *a* to *A* and from *b* to *B*. To implement it, we propose an algorithm of contour replacement.

Figures 2a and 2b present graphically the contour replacement. Within each elementary age group *i*, the procedure includes a sequence of replacements over the contours $B \rightarrow b \rightarrow a \rightarrow A$ (Figure 2a) and $A \rightarrow a \rightarrow b \rightarrow B$ (Figure 2b). Let us consider the former direction of replacement, which transforms vector m_B into vector m_A . The replacement starts from the youngest age group (i = 1). Following (1), the youngest age component of the cross-sectional difference at the second time point *T* is



Figure 2a. The *i*-th step of the contour decomposition trough transformation of vector *B* into vector *A*.



Figure 2b. The *i*-th step of the contour decomposition through transformation of vector *A* into vector *B*.

Using a simple algebraic trick, it is possible to express the component Δ_{AB}^1 differently:

$$\Delta_{AB}^{1} = \left[f\left(\boldsymbol{m}_{bB}^{[1]}\right) - f\left(\boldsymbol{m}_{B}\right) \right] + \left[f\left(\boldsymbol{m}_{aB}^{[1]}\right) - f\left(\boldsymbol{m}_{bB}^{[1]}\right) \right] + \left[f\left(\boldsymbol{m}_{AB}^{[1]}\right) - f\left(\boldsymbol{m}_{aB}^{[1]}\right) \right]$$
(5)

The second step begins from the vector *B* with its first element replaced by the first element of vector *A*. In general, for the *i*-th step equation (5) can be written as follows:

$$\Delta_{AB}^{i} = \left[f\left(\boldsymbol{m}_{BB}^{[i]}\right) - f\left(\boldsymbol{m}_{AB}^{[i-1]}\right) \right] + \left[f\left(\boldsymbol{m}_{aB}^{[i]}\right) - f\left(\boldsymbol{m}_{BB}^{[i]}\right) \right] + \left[f\left(\boldsymbol{m}_{AB}^{[i]}\right) - f\left(\boldsymbol{m}_{aB}^{[i]}\right) \right], \quad (6)$$

$$i = 2, 3, \dots n.$$

The latter equation corresponds to the sequence depicted in Figure 2a. Figure 3 further explains the sequence of replacements at the level of vectors' elements. Instead of the direct replacement of the first element of vector *B* by the first element of vector *A* as we do it conventionally (dashed arrow in Figure 3), we pass through vectors *a* and *b* (solid arrows). At the second step, the replacement sequence is repeated for the second elements of all vectors and so on.

In equation (6), the first and the third additive terms are the trend (within-country) components of the change. They are produced by mortality changes in populations *B* (former *b*) and *A* (former *a*) and are defined as

$$\delta_{bB|B}^{i} = f\left(\boldsymbol{m}_{bB}^{[i]}\right) - f\left(\boldsymbol{m}_{AB}^{[i-1]}\right), \ i = 1, \dots, n, \ m_{AB}^{[0]} = m_{B}$$
(7)

$$\delta_{Aa|B}^{i} = f\left(\boldsymbol{m}_{AB}^{[i]}\right) - f\left(\boldsymbol{m}_{aB}^{[i]}\right), \quad i = 1, \dots, n$$
(8)

The second additive term in (6) is an initial conditions (between-country) component of the change

$$\Delta_{ab|B}^{i} = f\left(\boldsymbol{m}_{aB}^{[i]}\right) - f\left(\boldsymbol{m}_{bB}^{[i]}\right), \quad i = 1, \dots, n$$
(9)



Figure 3. The sequence of element replacement in the four vectors in the direction $B \rightarrow b \rightarrow a \rightarrow A$.

An alternative sequence of replacements $A \rightarrow a \rightarrow b \rightarrow B$ can be expressed by

$$\Delta_{BA}^{i} = \left[f\left(\boldsymbol{m}_{aA}^{[i]}\right) - f\left(\boldsymbol{m}_{BA}^{[i-1]}\right) \right] + \left[f\left(\boldsymbol{m}_{bA}^{[i]}\right) - f\left(\boldsymbol{m}_{aA}^{[i]}\right) \right] + \left[f\left(\boldsymbol{m}_{BA}^{[i]}\right) - f\left(\boldsymbol{m}_{bA}^{[i]}\right) \right], \quad (6a)$$

$$i = 1, 2, \dots n.$$

The corresponding trend and initial conditions components of the change are

$$\delta_{aA|A}^{i} = f\left(\boldsymbol{m}_{aA}^{[i]}\right) - f\left(\boldsymbol{m}_{BA}^{[i-1]}\right), \ i = 1, ..., n, \ \boldsymbol{m}_{BA}^{[0]} = \boldsymbol{m}_{A},$$
(7a)

$$\delta_{Bb|A}^{i} = f\left(\boldsymbol{m}_{BA}^{[i]}\right) - f\left(\boldsymbol{m}_{bA}^{[i]}\right), \ i = 1, \dots, n ,$$
(8a)

$$\Delta_{ba|A}^{i} = f\left(\boldsymbol{m}_{bA}^{[i]}\right) - f\left(\boldsymbol{m}_{aA}^{[i]}\right), \ i = 1, \dots, n .$$
(9a)

The final trend components are determined by averaging the two equally possible sequences

$$\delta^i_{Aa|AB} = \frac{1}{2} \left[\delta^i_{Aa|B} - \delta^i_{aA|A} \right], \tag{10}$$

$$\delta^i_{Bb|AB} = \frac{1}{2} \left[\delta^i_{Bb|B} - \delta^i_{bB|A} \right] \tag{11}$$

The initial conditions component is the average of corresponding age-specific components

$$\Delta^{i}_{ab|AB} = \frac{1}{2} \left[\Delta^{i}_{ab|B} - \Delta^{i}_{ba|A} \right]$$
(12)

Equations (6)-(9), (6a)-(9a), and (10)-(12) fully determine the algorithm of contour replacement.

Equations (6), (6a) and (10)-(12) ensure that the equation

$$\Delta_{AB}^{i} = \Delta_{ab|AB}^{i} + \left(\delta_{Aa|AB}^{i} - \delta_{Bb|AB}^{i}\right), \quad i = 1, 2, \dots, n$$

is true.

R-program

R code to implement the decomposition of an inter-population difference in an aggregate demographic measure by age, initial conditions, and trend will be available as a Max Planck Institute for Demographic Research Technical Report.

Empirical examples

Example 1. Life expectancy gap between Japan and the USA in 2009 with a reference to 1970.

Using the Human Mortality Database (HMD, 2014) data, one can see an important increase in the difference in life expectancy between Japan and the USA from 1970 to 2009. Respective values of life expectancy at birth for men and women are shown in Table 1. While for men the initial difference of 2.3 years increased to 3.5 years, for women the negligible initial difference of 0.02 rose to 5.4 years.

		Males			Fomalos	
		IVIAICS			Ternales	
	USA	Japan	Diff.	USA	Japan	Diff.
1970	67.02	69.32	2.30	74.65	74.67	0.02
2009	76.13	79.61	3.48	81.04	86.42	5.38
Change	9.11	10.29	1.18	6.39	11.75	5.36

Table 1. Life expectancy in the USA and Japan in 1970 and 2009

For each sex, the contour replacement algorithm was applied to decompose the life expectancy difference between Japan and the USA in 2009 into age-specific initial conditions and age-specific trend components. The input data consisted of four vectors of age-specific death rates from the HMD (Japan in 1970 and 2009 and the USA in 1970 and 2009).

The decomposition results are displayed in Figure 4. The two upper panels display age-specific components of the difference and their initial conditions and trend parts. On the two lower panels the trend component is further split into parts produced by mortality change in Japan and those produced by mortality change in the USA.

The structure of the Japanese lifetime advantage in 2009 differed between men and women: For men the contributions peaked at middle adult ages and were smallest at younger and older ages, while for women old ages clearly dominated. In men 68% of the overall Japanese advantage in 2009 was produced by working ages 15 to 64 with the highest contribution by the age group 40 to 64. The impact of working ages was determined by the initial disadvantage of the USA, which was only slightly tempered by the effects of mortality change. Childhood ages also contributed somewhat to the Japanese advantage and these contributions were mostly determined by the initial conditions of 1970. By contrast, at ages 65 and older, the 2009 Japanese advantage owed exclusively to the trend component; American men had more favorable initial conditions over these ages.

For women, nearly 60% of the Japanese lifetime advantage in 2009 came from ages 65 and older; ages 65 to 79 produced the highest contribution followed by ages 80+. In both age groups the impressive trend component played the decisive role, which was somewhat tempered by an initial U.S. advantage. The contribution by the age group 40 to 64 was split into about equal parts due to mortality trends and initial between-country differences. One can see that in this age group mortality change made a positive contribution to the advantage of Japanese women. As we just mentioned, this was not the case for men.



Figure 4. Contributions to the total difference in life expectancy at birth between Japan and the USA in 2009 of age groups with further split into dynamic and initial (of the year 1970) conditions parts.

The two lower panels explain how the trend-age-components of the Japan-USA life expectancy gap of 2009 were formed. In all age groups, the U.S. experienced mortality decline which would have narrowed the life expectancy gap had the Japanese maintained their 1970 age structure of mortality. However, the Japanese mortality trends were stronger over most ages, especially for females, and thus the gap widened further.

Example 2. Difference in the age at first birth variability between Czech and Russian female cohorts of 1970 with a reference to 1955.

The left panel of Figure 5 presents age distributions of unconditional fertility rates over ages 12 to 39 for first births of female 1955 and 1970 birth cohorts in the Czech Republic and Russia (Human Fertility Database, 2014). For the 1970 cohort, the Czech and the Russian fertility rates were about the same at ages below 20 and in both distributions the mode corresponded to age 20. At ages 21 to 26 the Russian fertility rates were higher, but at ages above 26, the Czech rates were higher.

The Czech and Russian fertility rates compare quite differently for the older cohorts of 1955. Here Czech fertility was clearly higher below age 20 than Russian fertility, but lower at ages 21 and older. This suggests that between the cohorts of 1955 and 1970, Czech women experienced a massive redistribution of first births from ages 21 to 26 to ages 27 and older. In Russia, there was some fertility increase at ages below 20 and a similar (but weaker than in the Czech case) fertility redistribution from ages 21 to 27 to ages 27 and older.

Values of the standard deviation of the age at first birth are given in Table 2. In both countries the 1970 birth cohort had a higher standard deviation than the 1955 cohort. But the increase was greater in the Czech Republic than in Russia.

Using the contour decomposition method, we estimated age-contributions to the Czech-Russia difference between the 1970 cohort standard deviations with reference to the 1955 cohorts (Figure 5, right panel). As much as 77% of the total Czech-Russian difference in 1970 was produced by fertility differences at ages above 30. For this age group, the trend and initial conditions components were working against each other, with a stronger impact from the trend component. While the age group 12-19 contributed a little toward an increase in the Czech-Russia difference, ages 20 to 29 contributed somewhat in the opposite direction. At all age groups under 30, impacts of the trend were stronger compared to impacts of initial conditions.

Table 2. Standard deviation in the age at first birth over ages 12 and younger to 40 in the
Czech Republic and Russia in female cohorts of 1955 and 1970

	Standard deviation				
Cohorts	Czech Republic	Russia	Diff.		
1955	3.42	3.80	-0.38		
1970	4.55	4.11	0.44		
Change	1.13	0.31	0.82		



Figure 5. Unconditional age-specific fertility rates of first birth for female cohorts of 1955 and 1970 in the Czech Republic and Russia (left panel) and decomposition of the 1970 cohort difference in the standard deviation of first births into initial conditions (1955 birth cohort) and trend components (right panel).

Conclusion

When one thinks about the reasons for current differences between two populations in an aggregate measure of mortality, fertility or population reproduction, it is natural to be interested in the origin of this difference. To date, a scholar would normally have examined the components of between-population differences in the past and the changes between the starting period and the present or final time period. Information obtained from such analysis would include differences and changes attributed to each elementary event rate. Since these differences were numerous and often conflicting, their relative importance to the aggregate outcome measure remained unclear. The decomposition method proposed in this study permits a difference in an aggregate measure at a final time point to be split into additive components that correspond to the initial differences in the event-rates of the measure and differences in trends in these underlying event-rates.

In this study the aggregate measure was defined as a function of a vector of age-specific event rates. Technically our central task was to develop a decomposition method which ensured that the sum of the initial conditions and trend components equaled the conventional age-component (i.e. from other decomposition methods) of the inter-population difference at the final time point. This was accomplished by the contour replacement algorithm.

In the examples provided, we considered the aggregate measure to be a function of a onedimensional vector. It is certainly possible to include additional dimensions of interest in the decomposition such as causes of death for mortality or birth orders for fertility. In this case each step of the contour replacement of a single age-specific event rate would include a sequence of replacements of age-and-cause or age-and-birth order specific rates. This is analogous to the inclusion of additional dimensions within the framework of the general stepwise replacement algorithm as described in earlier studies (Andreev, Shkolnikov, Begun, 2002; Shkolnikov, Andreev, Zhang, Oeppen and Vaupel, 2011).

Acknowledgements

This work was completed as a part of the project on Divergence and Causes of Death (MODICOD) by the AXA Foundation and the project on Disparities in Mortality Trends to Future Health Challenges (DIMOCHA) by the Agence nationale de la recherché (ANR) and the Deutsche Forschungsgemeinschaft (DFG).

We are very grateful to Alyson van Raalte for editing the text and thoughtful suggestions. We would like to thank Evgeny Andreev for comments on the method and participants of the DIMOCHA-MODICOD seminar at the Institut national d'études démographiques (INED) for useful discussion on the 23rd of October 2014.

References

Andreev, E.M. 1982. Metod komponent v analize prodoljitelnosty zjizni. [The method of components in the analysis of length of life]. *Vestnik Statistiki*, 9, 42-47.

Andreev E.M., Shkolnikov V.M., 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*, 7: 499-522. Available at <u>www.demographic-research.org</u>

Arriaga E. 1984. Measuring and explaining the change in life expectancies. Demography 21(1), 83-96.

Beltran-Sanchez H., Preston S.H., Canudas-Romo V. 2008. An integrated approach to cause-ofdeath analysis: cause-deleted life tables and decomposition of life expectancy. *Demographic Research* 19: 1323-50. Available at <u>www.demographic-research.org</u>

Caswell, H. 2001. *Matrix Population Models: Construction, Analysis, and Interpretation* (2nd ed.). Sunderland, Massachusetts: Sinauer.

Horiuchi S., Wilmoth J.R., Pletcher S.D. 2008. A decomposition method based on a model of continuous change. *Demography* 45(4): 785-801.

Human Mortality Database. Available at www.mortality.org

Human Fertility Database. Available at www.humanfertility.org

Nau, C., & Firebaugh, G. 2012. A new method for determining why length of life is more unequal in some populations than in others. *Demography*, 49: 1207-1230.

Pressat, R. 1985. Contribution des écarts de mortalité par âge à la différance des vies moyennes. *Population*, 4-5, 766-770.

Shkolnikov V.M., Andreev E.M., Zhang Z., Oeppen J., J.W.Vaupel. 2011. Losses of expected lifetime in the United States and other developed countries: methods and empirical analyses. *Demography*, 48: 211-239.

Shkolnikov, V., Valkonen, T., Begun, A., Andreev, E. 2001. Measuring inter-group inequalities in length of life. Genus, LVII(3-4), 33-62.

van Raalte, A. A., & Caswell, H. 2013. Perturbation analysis of indices of lifespan variability. *Demography*, 50: 1615-1640.