

WP7: Decomposition tools
Technical report on decomposition

June 2013

Start date of project: 20 April 2012
Erasmus MC

Duration: 36 months
Draft 2

Table of contents

Abstract	4
Acknowledgements	5
Introduction	6
Part 1: The attribution method.....	7
1. Background of the decomposition method.....	7
2. Rationale of the attribution method.....	8
3. General principle of the decomposition method	8
4. Sullivan and Arriaga methods as core of the decomposition method	9
4.1 Sullivan method.....	9
4.2 Arriaga method.....	9
4.3 Decomposition method	10
5. Decomposition of life expectancy with disability into mortality vs. disability effect	10
6. Decomposition of life expectancy without disability into mortality vs. disability effect	11
6.2 Decomposition of mortality effect by cause	13
6.3 Decomposition disability effect by cause.....	14
6.4 Decomposition of total effect by cause	15
7. Discussion of the decomposition method.....	15
7. Discussion of the decomposition method.....	15
Part 2: The decomposition software tool.....	17
1. Introduction of the software tool.....	17
2. Features of the tool.....	18
2.1 General option 1: Family of decompositions	18
2.2 General option 2: Different types and formats of input data.....	19
2.2.2 Mortality data: life table by single-year of age.....	21
2.2.3 Mortality data: abbreviated life table	21
2.2.4 Disability prevalence (proportion) by cause, age and population.....	21
2.2.5 Disability prevalence (proportion) by age and population	21
2.2.5 Disability prevalence (proportion) by age and population.....	22
2.2.5 Disability prevalence (proportion) by age and population.....	22
2.3 General option 2: Decomposition of life and health expectancy from different starting ages.....	22
2.4 General option 3: Option to calculate and decompose partial life expectancies.....	22
2.4 General option 3: Option to calculate and decompose partial life expectancies.....	22
3. How to use the tool.....	23
3.1 Input preparations and running the model.....	23
3.1.1 Input dataset	23
3.1.2 The Excel input file.....	28
3.1.3 Running R.....	30
3.2 Output of the model.....	30
Part 3: Example	34
1. Input	34
1.1 Input dataset (ASCII or POR-format)	34
1.2 Input specification file (.csv).....	36

1.3 R syntax file (.R)	36
2. Output.....	37
3. Illustration of features of the tool.....	44
3.1 Illustration of different starting ages	44
3.2 Illustration of partial life expectancies	45
3.2 Illustration of partial life expectancies	45
Appendices	47
References	53

Abstract

This report focuses on this decomposition tool to decompose differences in health expectancy into the additive contribution of the mortality vs. disability effect, of different age groups and of different causes. The first part of the report is devoted to the method, which was previously published in *Demography* (Nusselder and Looman 2004). The second part focuses on the tool; it describes its main features, explains how to use the tool and provides some illustrations. This technical report is an updated version of the technical report of July 2011. The main additions as compared to the prior version are: 1) option to calculate confidence intervals based on bootstrapping and 2) possibility to start the analyses also from death and population numbers in 5-year age groups. The extended tool is available on request from the authors (w.nusselder@erasmusmc.nl).

Acknowledgements

This document is delivered as part WP9 of the JA EHLEIS. More information on JA-EHLEIS can be found on the website (<http://www.eurohex.eu/index.php?option=welcome>). The current work in the JA-EHLEIS builds upon the work in the EHEMU and EHLEIS project. In the EHEMU project (2004-2007) the decomposition method to decompose differences in health expectancy into the additive contribution of the mortality vs. disability effect, of different age groups and of different causes, was developed. In the EHEMU project (2007-2010) a tool was developed to decompose differences in health expectancy. In the current JA EHLEIS (2011-2014) the possibilities to derive confidence intervals and to start the analyses from death and population numbers in 5-year age groups were added.

This report was prepared by Wilma Nusselder and Caspar Looman. It reflects only its authors' views; the European Commission and those who provided information are therefore not liable for any use that may be made.

Introduction

Health expectancy is increasingly used to monitor population health. To explain differences in HE between populations (men vs. women, member states) and over time, differences in HE can be decomposed into the additive contribution of (1) differences in total mortality vs. differences in disability, (2) different age groups, and (3) different causes (diseases) (Nusselder and Looman 2004).

This is the first of two technical reports prepared within WP9 of the JA EHLEIS, focussing on the decomposition tool. The decomposition tool allows decomposing differences in health expectancy into the additive contribution of the mortality vs. disability effect, of different age groups and of different causes. But the tool can also be used to decompose differences in total life expectancy by age and/or cause. Depending on available data, different decompositions can be made (“family of decompositions”).

Application of the tool to decompose differences in health expectancy requires standard data to calculate health expectancy, i.e. a life table or population and death data to build a life table, and age-specific prevalence of disability. Apart from this requirement, the tool is flexible. It allows the user to work with different types and formats of input data (txt and xls) and has the option to include causes of death and disability. The user is free to select diseases and age categories in the analyses. The user also has the option to obtain additional output on the life table or on the life table decomposition. A new feature is that confidence intervals based on bootstrapping can be obtained. The tool is programmed in R, but the user does not need to have any R knowledge to run the program.

The decomposition and attribution tool are each described in a separate technical report, which can be read independently. The tools are freely available from the author.

The current document on the decomposition tool is organised as follows. In part 1, we give the background, rationale and the general principle of the decomposition method, published in Demography (Nusselder and Looman 2004). In Part 2 we describe the tool that is based on this method, its main features, and explain how to use it. In Part 3 we illustrate the different types of decompositions that are possible with the tool. To allow independent reading of the different parts, some information is given in more than one part of the report.

Part 1: The decomposition method

1. Background of the decomposition method

Health expectancy indicators, that combine mortality and morbidity data into a single composite indicator, are increasingly used to monitor the health of populations and to compare health over time or between populations or population groups. Health expectancy indicators share important attractive properties with life expectancy, such as their measurement in expected years of life and their independence from the age structure of the population, if small age intervals are used. An additional attractive feature is that health expectancy takes into account both mortality and the health status of the surviving population and thus provides information on the length of life (adding years to life), and the healthfulness of life (adding life to years). A well-known example is the disability-free life expectancy (DFLE). DFLE indicates how many years of the total life expectancy a person of a given age can expect to live without disability, and the difference between the total life expectancy and the DFLE is the life expectancy with disability. Different variants of DFLE exist, depending on how disability is measured.

Next to describing differences in health expectancy, it is important to understand how and why the health of a population changes over time and why differences in health exist between population groups. Knowing which age groups and diseases contribute most to described differences in population health can point at potential determinants and can assist in the evaluation of past trends and health inequalities. This may facilitate the definition of priorities in the field of public health, and may improve the assessment of targeted health priorities.

In mortality research, decomposition (i.e. partitioning) techniques were developed in the 1980s to assess the contribution that age groups make to differences in life expectancy by Andreev (Andreev, Shkolnikov et al. 2002), Arriaga (Arriaga 1984) and Pressat (Pressat 1985). A continuous version of the method for decomposition of differences between life expectancies by age was developed by Pollard in 1982 (Pollard 1988). A general algorithm for the decomposition of differences between two values of an aggregate demographic measure, including life expectancies, was developed by Andreev and colleagues (Andreev, Shkolnikov et al. 2002). Decomposition tools have been widely used to explain differences in the length of life between sexes, socio-economic groups and time.

Decomposition tools are now available to partition differences in health expectancy. A method for decomposition of a difference between two health expectancies was independently developed in the early 2000s by Andreev (Andreev, Shkolnikov et al. 2002) and Nusselder (Nusselder and Looman 2004). Both methods are based on the Sullivan method and decompose health expectancy into the additive contributions of mortality vs. disability differences and of age groups. The method of Nusselder additionally decomposes differences in health expectancy into the additive contribution of causes of death and causes of disability.

The tool presented in this report is based on the publication of Nusselder and Looman in *Demography* (Nusselder and Looman 2004).

2. Rationale of the decomposition method

The technique to decompose differences (or changes) in health expectancy earlier published in *Demography* (Nusselder and Looman 2004) is based on the Sullivan method (Sullivan 1971) and is an extension of the decomposition method for life expectancy developed by Arriaga (Arriaga 1984; Arriaga 1989). The Sullivan method is the standard method to calculate health expectancy on a routine basis as it uses data from cross-sectional survey in combination with mortality data or life tables that are widely available. The Arriaga method is frequently used to decompose differences in life expectancy. While Arriaga's formula is written in a slightly different form, it is essentially equivalent to the formulae by Andreev and Pressat (Shkolnikov, Andreev et al. 2006).

3. General principle of the decomposition method

The general principle is first to decompose differences in years with (or without) disability into the contribution of mortality differences (mortality effect) vs. disability differences (disability effect). The mortality effect is then further decomposed by age and cause of death using a modification of the Arriaga method. This method assumes that within each age group, the contribution that a cause of death makes to the change in life expectancy between time t and $t+n$ is proportional to the contribution that this cause makes to the change in the central mortality rate in that age group.

The disability effect is further decomposed by cause of disability by using disability prevalence by cause (attributions), which can be obtained from cross-sectional data using the attribution tool. For more details on this tool and assumptions, see the technical report on the attribution tool (Nusselder and Looman 2010)

4. Sullivan and Arriaga methods as core of the decomposition method

4.1 Sullivan method

The Sullivan method is a prevalence-based life table technique to calculate health expectancy (Sullivan 1971). According to this technique, the number of person years in the age interval $x, x+i$ (${}_iL_x$) is subdivided into years with and without disability, by multiplying ${}_iL_x$ by the observed age-specific disability prevalence (i.e., the proportion with disability between age x and $x+i$, ${}_i\pi_x$). Summation of the number of person years with(out) disability across age gives the number of person years-lived beyond age a (T_a) with(out) disability. Dividing T_a with(out) disability by the exact number of survivors at age a (l_a) gives life expectancy with(out) disability at age a . The last step can be omitted in a life table with radix 1 ($l_0=1$). For more details on the Sullivan method to calculate health expectancy, see the Sullivan calculation guide (Jagger 1999) and examples on Eurohex-website (<http://www.eurohex.eu/index.php?option=welcome>.)

4.2 Arriaga method

The Arriaga method first decomposes the difference or change in life expectancy by age. For the sake of transparency, we explain the method in terms of change, hence a difference between time points, but the user can also read a difference between populations or sub populations.

Arriaga distinguishes three different effects of mortality changes on life expectancy: a direct effect (*DE*), an indirect effect (*IE*) and an interaction effect (*I*). The direct effect is the change in the number of person years lived within a particular age group (${}_iL_x$) as a consequence of a mortality change in that age group. The indirect effect is the number of years added to (or removed from) a given life expectancy, because a mortality change within a specific age group produces a change in the number of survivors at the end of that age interval. In the presence of unchanged mortality rates at older ages than the age group under consideration, the increase (or decrease) in the number of survivors at the end of the age interval results in an increase (or decrease) in the number of years lived. Both the direct and indirect effect take into account mortality change in a specific age group, independent of the changes in other ages. Since in general mortality changes occur simultaneously in all ages, in addition a small part of the change in life expectancy is due to the fact that the additional (or fewer) survivors (those responsible for the indirect effect) do not experience unchanged mortality at older ages. The effect resulting from the combination of the changed number of survivors at the end of the age interval and the lower (or higher) mortality rates at older ages is termed the interaction effect (*I*) (Arriaga 1984). Adding the direct, indirect and interaction effect gives the total contribution of each age group to the change in life expectancy, or in other words, the decomposition of a change in life expectancy by age. In the second step, the contribution of each age group is further decomposed by cause of death,

assuming that within each age group, the contribution that a cause of death makes to the change in life expectancy between time t and $t+n$ is proportional to the contribution that this cause makes to the change in the central mortality rate in that age group. The equations of the Arriaga method are given in the publication of Nusselder and Looman (Nusselder and Looman 2004), appendix A.

4.3 Arriaga decomposition method applied to Sullivan life table

While the decomposition method to decompose differences in health expectancy is based on the Arriaga method, to apply the Arriaga method to a Sullivan life table instead of a standard (single-decrement) life table for the calculation of life expectancy, additional steps are needed. Whereas changes in life expectancy are caused by changes in mortality rates only, changes in health expectancy (calculated with the Sullivan method) reflect changes in mortality rates and/or changes in the prevalence of disability. The first extra step is the decomposition into the contribution made by each of these components (i.e. the mortality and disability change), which we refer to as decomposition by type of effect.

Next, similar to the Arriaga method, mortality changes are decomposed by age, and finally causes of death and disability are incorporated. Whereas the Arriaga method decomposes the change in life expectancy by age (Arriaga 1984) and further by cause (Arriaga 1989), here we decompose the change in the number of person years lived *in each age interval* by age and cause. This modification is needed, because in the Sullivan method the proportion with disability in each age group is multiplied with the number of person years in that age group. For this reason, we re-expressed the Arriaga method. First, we made a distinction between the age group where the mortality change occurs ('age at origin' labeled as $y, y+i$) and the age group where the person years are added to, or removed from ('age at destination, labeled as $x, x+i$). Second, we re-expressed the number of person years lived after age x (T_x) in terms of ${}_iL_x$ (i.e., age-specific contribution to T_x), using that summation of ${}_iL_x$ over age gives T_x

5. Decomposition of health expectancy into mortality vs. disability effect

Core of the Sullivan life table is the number of person years with disability (${}_i\pi_x L_x$), in age group, $x, x+i$, (where i is the length of the age interval). This is the product of the number of person years lived (${}_iL_x$) and the proportion with disability (${}_i\pi_x$). A change in the number of person years with disability is:

$$\Delta_i \pi_x L_x = \left[\frac{{}_i\pi_{x(t)} + {}_i\pi_{x(t+n)}}{2} \right] \cdot \Delta_i L_x + \Delta_i \pi_x \cdot \left[\frac{{}_iL_{x(t)} + {}_iL_{x(t+n)}}{2} \right] \quad (1)$$

The change in the number of person years with disability is the sum of two components:

$${}_iMOR_x = \left[\frac{{}_i\pi_{x(t)} + {}_i\pi_{x(t+n)}}{2} \right] \cdot \Delta_i L_x \quad (2)$$

$${}_iDIS_x = \left[\frac{L_{x(t)} + L_{x(t+n)}}{2} \right] \cdot \Delta_i \pi_x \quad (3)$$

The first component, the mortality effect “ ${}_iMOR_x$ ” is the change in the number of person years with disability due to a change in the number of person years lived. The mortality effect is the change in the number of person years with disability that would occur if only mortality rates would change. A negative ${}_iMOR_x$, for instance, reflects a decline in the number of person years lived with disability in the age group $x, x+i$ due to an increase in the mortality in that age group, or in younger age groups. The second component, the ‘disability effect’ (${}_iDIS_x$), is the change in the number of person years with disability due to a change in the proportion with disability (*ceteris paribus*). The disability effect in one age group is the change in the number of person years with disability that would occur if only the proportion with disability would change. A negative ${}_iDIS_x$ reflects a decline in the number of person years lived with disability in the age group $x, x+i$ due to a decline in the proportion with disability in that age group.

For the decomposition of the change in the number of years without disability (DFLE), the approach is similar. The only difference is that the proportion without disability (i.e., $1 - {}_i\pi_x$) rather than the proportion with disability (${}_i\pi_x$) is used in the equations.

6. Further decomposition of mortality and disability effect by age and cause

To decompose the mortality effect (${}_iMOR_x$, see equation 2) by age and cause, the change in the number of person years lived ($\Delta_i L_x$), is first decomposed into the contribution made by age and next by specific causes of death.

The disability effect is (equation 3) is already by age. Age-specific differences in prevalence of disability originate from differences in incidence, recovery and mortality rates at younger ages, however, based on cross-sectional prevalence data, these underlying dynamics cannot be further assessed.

Again we explain the decomposition for the life expectancy with disability. For the life expectancy without disability (DFLE), the proportion without disability (i.e., $1 - {}_i\pi_x$) rather than the proportion with disability (${}_i\pi_x$) is used in the equations.

6.1 Decomposition of mortality effect by age

Similar to the Arriaga method, we define the direct, indirect, interaction and total effect, though we distinguish between age groups of origin y , $y+i$ and the age groups of destination x , $x+i$. The direct effect (${}_iDE_{xy}$) of a mortality change in the age group y , $y+i$ between time t and $t+n$ on the number of person years lived between age x , $x+i$ is expressed as follows:

$${}_iDE_{xy} = \frac{l_y^t}{l_a^t} \cdot \left[\frac{{}_iL_x^{t+n}}{l_y^{t+n}} - \frac{{}_iL_x^t}{l_y^t} \right] \quad (x=y) \quad (4a)$$

$${}_iDE_{xy} = 0 \quad (x > y) \quad (4b)$$

The formula for the last open-ended age group is as follows:

$$DE_{xy} = \frac{l_y^t}{l_a^t} \cdot \left[\frac{T_x^{t+n}}{l_y^{t+n}} - \frac{T_x^t}{l_y^t} \right] \quad (y=x, \text{ open ended age group}) \quad (5)$$

A mortality change in the open-ended age group causes only a direct effect. For the other age groups in addition to the direct effect also the indirect and interaction effect have to be calculated.

The indirect effect consists of the number of person years added (or removed) because the mortality change within a specific age group of origin produces a change in the number of survivors at the end of that age interval. The indirect effect is the effect that would arise if the changed number of survivors would continue living after age $y+i$ as many years as the rest of the population before the change in mortality (i.e. the life expectancy at age $y+i$ before the change in mortality). The formula for the indirect effect ${}_iIE_{xy}$ is:

$${}_iIE_{xy} = \frac{{}_iL_x^t}{l_a^t} \cdot \left[\frac{l_y^t \cdot l_{y+i}^{t+n}}{l_{y+i}^t \cdot l_y^{t+n}} - 1 \right] \quad (x > y) \quad (6)$$

The interaction effect (${}_iIE_{xy}$) is calculated as the difference between two components: (1) the number of person years added (removed) if the additional survivors at age $y+i$ would continue living as many years as the rest of the population after the change in mortality (i.e. the life expectancy at age $y+i$ after the change in mortality) and (2) the indirect effect, being the number of years added (removed) if the additional survivors would continue living under the old mortality regime. The first component labeled ${}_iOE_{xy}$ is calculated as follows:

$${}_i OE_{xy} = \frac{{}_i L_x^{t+n}}{l_a^t} \cdot \left[\frac{l_y^t}{l_y^{t+n}} - \frac{l_{y+i}^t}{l_{y+i}^{t+n}} \right] \quad (x>y) \quad (7)$$

Using equation (8) and (9), the interaction effect (${}_i I_{xy}$) is:

$${}_i I_{xy} = {}_i OE_{xy} - {}_i IE_{xy} \quad (x>y) \quad (8)$$

The total contribution of a mortality change in each age group of origin $y, y+i$, to the change in the number of person years lived between age x and $x+i$ ($\Delta {}_i L_x$) is calculated as follows:

$${}_i TOT_{xy} = {}_i DE_{xy} + {}_i IE_{xy} + {}_i I_{xy} \quad (x>y) \quad (9)$$

The total effect by age of origin $y, y+i$ gives this information the decomposition by age.

6.2 Decomposition of mortality effect by cause

For the decomposition of the total effect (${}_i TOT_{xy}$) by cause of death k , we follow the Arriaga method for the decomposition by cause of death (Arriaga 1989). It is assumed that the contribution of the change in mortality from each cause in age group $y, y+i$ to the change in the number of person years in age group $x, x+i$ (or: total life expectancy in the Arriaga method), is proportional to the contribution of this cause to the change in the central mortality rate in the age group $y, y+i$. We multiply ${}_i TOT_{xy}$ with the contribution of the mortality change in the age group $y, y+i$ attributable to cause k , ${}_i C_{yk}$, to obtain the decomposition by cause:

$${}_i TOT_{xyk} = {}_i TOT_{xy} \cdot {}_i C_{yk} \quad (x > y) \quad (10)$$

where ${}_i TOT_{xyk}$ is the contribution of a mortality change due to cause k in age group $y, y+i$ to the number of person years lived between age $x, x+i$ (where $y \leq x$). ${}_i C_{yk}$ is calculated using:

$${}_i C_{yk} = \left[\frac{R_{yk}^{t+n} \cdot {}_i M_y^{t+n} - R_{yk}^t \cdot {}_i M_y^t}{{}_i M_y^{t+n} - {}_i M_y^t} \right] \quad (11)$$

where ${}_i M_y^t$ is the central mortality rate at age $y, y+i$ at time t , ${}_i R_{yk}^t$ is the proportion of deaths from cause k in the total number of deaths in the age group $y, y+i$ at time t and n is the difference between the first year of observation and the second. The contribution of cause k to the change in the number of person years lived between age $x, x+i$ (${}_i TOT_{xk}$) is obtained as follows:

$${}_i TOT_{xk} = \sum_{y=a, x} {}_i TOT_{xyk} \quad (12)$$

The mortality effect by cause (${}_iMOR_{xk}$) is obtained as follows:

$${}_iMOR_{xk} = \left[\frac{{}_i\pi_{x(t)} + {}_i\pi_{x(t+n)}}{2} \right] \cdot \Delta_i L_{xk} = \left[\frac{{}_i\pi_{x(t)} + {}_i\pi_{x(t+n)}}{2} \right] \cdot TOT_{xk} \quad (13)$$

Summation of ${}_iMOR_{xk}$ over age x gives the total mortality effect by cause, i.e. the change in health expectancy that would occur if only mortality from that specific cause k would have changed.

In order to avoid that the results of the decomposition depend on whether the first or second time point is used as reference, we averaged the components of the difference between time t and $t+n$ (with $t+n$ as baseline) with the respective components of the difference between population $t+n$ and t (with t as baseline). This additional step was proposed by Andreev and Pressat (Shkolnikov et al. 2001); the original Arriaga method did not include this step.

6.3 Decomposition disability effect by cause

For each age group, in order to decompose the disability effect (${}_iDIS_x$) by cause of disability, the change in the proportion with disability ($\Delta_i\pi_x$) needs to be attributed to different causes of disability k . Disability by cause can either be obtained from disability surveys, or can be estimated using individual-level data on chronic diseases and disability from health surveys using the attribution tool (Nusselder and Looman 2010). Similar to causes of death, causes of disability (attributions) should be additive, that is, the sum of all causes should equal total disability. The change in the proportion with disability decomposed by cause can be obtained by subtraction of the proportion of disability for each cause k at time t from that at time $t+n$, or:

$$\Delta_i\pi_{xk} = {}_i\pi_{xk}^{t+n} - {}_i\pi_{xk}^t \quad (14)$$

where ${}_i\pi_{xk}^{t+n}$ is the proportion with disability from cause k in age group x , $x+i$ at time $t+n$, and ${}_i\pi_{xk}^t$ is the proportion with disability from cause k in age group x , $x+i$ at time t . And where:

$$\Delta_i\pi_x = \sum_k \Delta_i\pi_{xk} \quad (15)$$

Substitution of the change in the proportion with disability by cause in equation 5 gives the disability effect by cause (${}_iDIS_{xk}$):

$${}_iDIS_{xk} = \left[\frac{{}_iL_{x(1)} + {}_iL_{x(2)}}{2} \right] \cdot \Delta_i\pi_{xk} \quad (16)$$

Summation of ${}_iDIS_{xk}$ over x gives the total disability effect by cause, i.e. the change in health expectancy that would occur if only disability from that specific cause k would have changed.

6.4 Decomposition of total effect by cause

Summation of the mortality and disability effect by cause gives the decomposition of the total change in health expectancy by cause.

7. Calculation of confidence intervals

Parametric bootstrapping was used to obtain confidence intervals (Efron and Tibshirani 1994). Bootstrapping, sometimes called re-sampling, constructs a new dataset (“replica”) which contains a random sample of the original observations with replacement. This mimics the situation where you would have repeated the collection of your data. This leads to different estimates of the parameters. If this is repeated for instance 1000 times it gives an impression of the uncertainty of the parameters (disabling impacts).

Non-parametric bootstrap makes no assumptions about the distributional form of the population or probability (parent) distribution. However, when it is known from which family of distributions the parent distribution belongs to, parametric bootstrapping can be used. For the parametric bootstrap, one selects the distribution type one believes the data to come from and then find the parameters for that distribution. In case of parametric bootstrapping, each observation is replaced with a “sample” taken at random from the fitted population distribution. In this way a bootstrap sample “replica” is generated, which has the same (sample) size as the original data. Then the outcomes (parameters) for which confidence intervals are needed are calculated from the bootstrap sample. This sampling process is repeated many times, yielding a distribution of each outcome.

In the decomposition analysis, mortality rates (derived from number of deaths and mid-year population, total or by cause) and the prevalence of disability (number of disabled as fraction of total number). The number of deaths is generally considered to be Poisson distributed and the prevalence binomially distributed.

Comment [DW1]: Wat is er met de mortality rate en prevalence of disability?

8. Discussion of the decomposition method

A few issues may merit extra explanation or discussion.

The decomposition method is based on the Sullivan method, because this is the standard method to calculate health expectancy on a routine basis. Health expectancy derived from the Sullivan method reflects the current health composition of a real population adjusted for mortality. The major

disadvantage of the Sullivan method is that it integrates observed prevalence data on disability (i.e. stock data) in a life table which is based on incident data on mortality (i.e. flow data). As a result the Sullivan method generally does not produce a pure period indicator such as (period) life expectancy (Barendregt, Bonneux et al. 1994; Van de Water, Boshuizen et al. 1995; Mathers and Robine 1997) The deviation from a pure period indicator can introduce bias. A change in health expectancy may have been caused by prevalence reaching its equilibrium values, instead of a real change in population health. Consensus seems to exist that serious biases are only expected to occur when sudden, large changes (Barendregt, Bonneux et al. 1994; Van de Water, Boshuizen et al. 1995; Mathers and Robine 1997). Nonetheless, conclusions about changes in health expectancy assessed with the Sullivan method require some caution. Less concern is warranted for comparisons between population groups (such as men and women) in a period without large changes (such as in the Netherlands in the 1990s).

That the decomposition method is based on the Sullivan method has two consequences for the interpretation of the outcomes. First, the decomposition method quantifies to what extent differences in disability prevalence and total mortality (in each age group) contribute to health expectancy differences, but does not indicate to what extent these differences are due to differences in age-specific transition rates into and out the disabled state and mortality rates from the nondisabled and disabled state. Second, the disability effect by age shows to what extent differences in the prevalence of disability in specific age groups contribute to the difference in health expectancy. Prevalence of disability in each age group is the net result of transitions in and out of disability and into death from birth up to that and including that age group. These underlying dynamics cannot be further assessed, and the disability effects by age points where age-specific differences in prevalence occur, but not where they originate.

Part 2: The decomposition software tool

1. Introduction of the software tool

The decomposition tool allows the user to decompose differences in health expectancy differences into the additive contribution of mortality vs. disability effects, of different age groups and of different causes. The software tool implements the decomposition method as published by Nusselder and Looman in *Demography* (Nusselder and Looman 2004). This decomposition method is based on the Sullivan method to calculate health expectancy and is an extension of the Arriaga method (Arriaga 1984) to decompose differences in total life expectancy. The decomposition tool first decomposes the difference (or change) in health expectancy at certain age into mortality and disability effects. This is called “decomposition by type of effect”. The mortality effect is the difference in health expectancy due to a difference in the number of person years lived caused by a difference in total mortality up to and including that age. The disability effect is the difference in health expectancy due to a change in the proportion with disability. The mortality effect is further decomposed by age of origin using an adjustment of the Arriaga method. If data on disability and mortality by cause are included, the tool allows further decomposition by cause of death and disability.

The decomposition is developed to decompose differences in health expectancy by type of effect, age and cause, but the tool allows different decompositions. These range from decomposition of life expectancy by age, using only mortality data by age and sex, to the full decomposition of health expectancy by type of effect, age and causes, using mortality and disability data by age, sex and cause.

The tool decomposes differences in health or life expectancy between two populations. Two populations can refer to different populations (e.g. member states), different sub-populations (e.g. gender, socio-economic groups) but also to different time periods.

The decomposition tool is flexible as it allows the user to work with different types and formats of input data, both for the mortality and the disability data. The user can obtain detailed outcomes on the life table calculations and the Arriaga decomposition method. The decomposition software is programmed in R, but all the user-specified input is communicated to the R program through an Excel-file (saved as “csv-file”), therefore user does not need to have any R knowledge to use the program.

Before we explain in more detail how the tool works, we first explain the features of the tool.

2. Features of the tool

The decomposition tool has the following features. First, the tool allows making different decompositions. Second, the tool offers the possibility to use different types of input data, both for the mortality data and the disability data, allowing using existing databases maximally. Third, the tool offers the possibility to decompose life and health expectancy from different starting ages. Fourth, it allows the decomposition of partial life and health expectancies.

2.1 General option 1: Family of decompositions

The tool allows different decompositions. Table 1 presents the possible decompositions, depending on whether or not data are available on causes of death, on prevalence of disability, and on prevalence of disability by cause (attributions).

Table 1 Family of decompositions

	LIFETABLE data: Always	
	Causes of death	
Disability	No	Yes
No information	1.LE difference by age	2.LE difference by cause Also - LE difference by age
Prevalence by age	3.HE difference by type of effect Also: - LE difference by age - Disability effect by age - Mortality effect by age	4.HE difference: mortality effect by cause of death Also: - LE difference by age - LE difference by cause - HE difference by type of effect - Disability effect by age - Mortality effect by age
Prevalence by cause (attributions)	5.HE difference: disability effect by cause of disability Also: - LE difference by age - HE difference by type of effect - Disability effect by age - Mortality effect by age LE with disability by cause (of disability)	6.HE difference by cause (death and disability) Also - LE difference by age - LE difference by cause - HE difference by type of effect - Disability effect by age - Mortality effect by age - Disability effect by cause - Mortality effect by cause - Total effect by cause - LE with disability by cause of disability

2.2 General option 2: Different types and formats of input data

The tool offers the possibility to use different types of input data.

For the life table data four options exist:

1. Number of deaths and population at risk by single-year of age and population (variant DPs)
2. Number of deaths and population at risk by five-year of age and population (variant DPa, new)
3. Life table by single year of age and population (variant LTs)
4. Abridged life table (5 year or larger groups) by population (variant LTa)

For disability data, the following options exist:

1. Disability prevalence by age, population and causes (attributions)
2. Disability prevalence (proportion) by age (single year of age, “as” or age groups, “ag”) and population
3. No disability data (only decomposition of total life expectancies)

For mortality data, the following options exist:

1. Causes of death (numbers) by age groups and population
2. No causes of death

The family of decompositions, in combination with different data choices, yield a large number of variants for which the decomposition tool can be used. Table 2 presents the names we use for each variant in this technical report.

The tool can read life table data and disability prevalence by age in txt format (easy to obtain from the Human Mortality Database) and Excel format (easy to obtain from EHEMU database). Attributions must be a text file; all other files must be in Excel format.

Table 2 Naming of decompositions (jobs)

Variant	DPa	DPs	LTs	LTa
1	input decomposition 1-DPa.csv	input decomposition 1-DPs.csv	input decomposition 1-LTs.csv	input decomposition 1-LTa
2	input decomposition 2-DPa.csv	input decomposition 2-DPs.csv	input decomposition 2-LTs.csv	input decomposition 2-LTa
3	input decomposition 3-DPa-as.csv input decomposition 3-DPa-ag.csv	input decomposition 3-DPs-as.csv input decomposition 3-DPs-ag.csv	input decomposition 3-LTs as.csv input decomposition 3-LTs ag.csv	input decomposition 3-LTa
4	input decomposition 4-DPa-as.csv input decomposition 4-DPa-ag.csv	input decomposition 4-DPs-as.csv input decomposition 4-DPs-ag.csv	input decomposition 4-LTs as.csv input decomposition 4-LTs ag.csv	input decomposition 4-LTa
5	input decomposition 5-DPa-cause	Input decomposition 5-DPs-cause	input decomposition 5-LTs cause.csv	input decomposition 5-LTa
6	input decomposition 6-DPa-cause	Input decomposition 6-DPs-cause	input decomposition 6-LTs cause.csv	input decomposition 6-LTa

DPs=deaths or person years by single year of age
 DPa=deaths or person years by age group
 LTs=life table by single year of age
 LTa=life table by age groups: abbreviated or abridged life table
 ag=disability by age group
 as = disability by single year of age
 cause= disability by cause

2.2.1 Life table data: Number of deaths and population at risk

The first variant of data to construct the life table is based on the number of deaths and population at risk, by single year of age of five year age groups. For the population at risks the mid-year population can be used, or the exposure to risk population. The mid-year population is simply obtained by taking the average of the population (at a given age) in two adjacent years (e.g. in combination with deaths 2006, the mid-year population is the average of the population per January 1st 2006 and per January 1st 2007). Information on the number of deaths and population per January 1st by sex and single year of age is available from EHEMU database (<http://www.chemu.eu/>) or the Human Mortality database (<http://www.mortality.org/>). The Human Mortality database also provides the population at risk. We use in this document and in the file names the abbreviation “DP” for this variant, where D stands for deaths and P for population.

Based on the number of deaths and population at risk, a life table is built according to the Sullivan calculation guide (Jagger 1999). A simplification in the variant starting from the number of deaths and population at risk is that deaths are assumed to occur in the middle of the age interval, also in the age group 0-1 years. Life and health expectancies at birth based on this variant may therefore be slightly different from the Eurostat method, that also uses information on the number of births (Jagger 1999).

To avoid this, the user can start from an existing life table (see next section) or specify “ax” the fraction of the year alive. If no ax is provided it is set to 0.5.

2.2.2 Life table data: life table by single year of age

The tool can also start from a published single-year life table. This assures that the decomposed life or health expectancy is identical to a published one, and hence avoids slightly different outcomes because of different assumptions in the life table. Also this is practical for users who made already a (Sullivan) life table themselves and want to decompose health or life expectancy based on this life table.

However, with this variant bootstrapping to obtain confidence intervals is not possible.

To start from the life table, the user should provide information on age, lx (number of survivors at age x) and dx (number of death between age x and x+1) and the population.

2.2.3 Life table data: abridged life table

The tool can also start from a published abridged life table. An abridged life table includes age groups, in stead of single year ages. Similar to the variant starting from a life table by single year of age, this assures that the decomposed life or health expectancy is identical to an already published life or health expectancy or a life table already made in Excel. To start from the abridged life table, the user should provide information on age, lx (number of survivors at age x) and dx (number of death between age x and x+1). The tool requires that the age classes in the abridged life table are the same as those used in the file with mortality data by age and cause and in the file with the disability prevalence data. We advise to use 5-year age groups. Larger age groups can be used, but decrease the accurateness of the life table. If the user wants to use larger age intervals, mortality by cause of death should be classified into these larger age intervals as well. If no information on causes of death is used, the user should provide in the input specification file the cut-levels of the age categories, also for the situation of 5-year age groups. The reason that the user has to specify age groups also in the situation of 5-year age groups is that age intervals may be different for the last age group, for instance 85+ vs. 85-89 and 90+.

2.2.4 Disability prevalence (proportion) by age, population and cause

Disability prevalence by cause can be derived with the attribution tool (Nusselder and Looman 2010). The attribution tool generates a file (with extension mrc, or in case of bootstrapping with mrcxx) that can serve as input to the decomposition tool.

2.2.5 Disability prevalence (proportion) by age and population

In case disability prevalence by cause is not available, the tool can be used with disability prevalence by age to decompose health expectancy into the additive contribution of the mortality vs. disability

effect and of ages. In combination with single year life table, disability prevalence by single year of age (referred to as “S”) or by age group (referred to as “A”) can be used. In combination with an abridged life table only prevalence of disability by age groups can be used.

2.2.6 No information on disability

In case no information on disability is available, the tool can still be used to decompose total life expectancy by age, and if information on mortality by cause is present, by cause of death.

2.2.7 Causes of death by age and population

Data on causes of death are generally available, for instance from the WHO mortality database (<http://www.who.int/whosis/mort/download/en/index.html>). Causes of death can be entered as numbers or as proportions.

2.3 General option 3: Decomposition of life and health expectancy from different starting ages

The tool can calculate and decompose health and life expectancies from any starting age, as long as this starting age is a cut-off point in the life table and all data are available for the included ages. Thus for a life table with age classes, 5-9, 10-14, the life table can start at age 5, 10, etc. but not 12. When information for causes of death and/or disability, if included in the analysis, is available only above a certain age, the user should select this or a higher age to start the life table.

2.4 General option 4: Partial life expectancies

The tool can calculate and decompose partial health and/or life expectancies. A partial life expectancy is a life expectancy within a specified age range, for instance, life expectancy between age 15 and 80 years of age.

2.5 General option 4: Confidence intervals using bootstrapping

The tool can calculate confidence intervals using bootstrapping for the variants that start from number of deaths and the population at risk. For variants including disability prevalence (ag or as in variant 3 or 4), the number of cases in the survey should be provided by the user. In case information on disability by cause is used (cause in variant 5 or 6) the attribution tool should also be run with bootstrapping with the same number of bootstraps so that an mrcxx file is made with all the attribution tables for all the bootstrap samples.

3. How to use the tool

The decomposition tool is programmed in R. All the user-specified input is communicated to the R program through the input specification Excel-file (saved as “csv-file”). Hence, the user does not need to have any R knowledge to use the program.

To use the decomposition tool in R you need:

- I. Input datasets with life table or mortality data, disability data (optional) and cause of death data (optional)
- II. Input specification file (csv format or txt format)
- III. R syntax: “decomposition.R”

If R is not already installed, do first the following:

- Go to <http://lib.stat.cmu.edu/R/CRAN/>
- Choose the system you are using, probably ‘Windows(95 and later)’
- Click ‘base’
- Choose: “Other builds”
- Choose previous releases
- Choose version R-2.7.1

Further follow the instructions.

Having R installed on the computer, do the following:

- Put the script and data files in one folder.
- Check that the separator is a comma (check the computer), via start\settings\control panel\regional options.

3.1 Input preparations and running the model

The user only has to specify in the R syntax the path and name of the Excel input specification file (csv-file). As an alternative to the Excel input specification file, a text input specification file can be used, though this is not the preferred option. The input specification file has to include the name and location of the data files with the life table data, disability data (optional) and cause of death data (optional). Additionally this input specification file includes all the user choices for the options provided. Running a specific input specification file results in an analysis and output-file. This will be further named as “job”. The output is written to files (text format), which can be read by e.g., TextPad, Notepad, or MS-Word. Users can opt for additional output.

3.1.1 Input datasets

Here follows a description of the input datasets for each variant of the decomposition tool (Table 3). All data need to be provided by population, as the decomposition tool always decomposes differences between two populations. A population can be a country, but also one gender, or population for specific calendar year or period.

Table 3 Overview of data needed for each variant

INFORMATION TO CONSTRUCT THE LIFE TABLE		
All variants	INFORMATION TO CONSTRUCT THE LIFE TABLE Population and deaths by single year (DPs) or age groups (DPa), or life table by single year (LTs) or by age groups (LTa)	
Additional information	DISABILITY	CAUSE OF DEATH
Variant 1	-	-
Variant 2	-	Number of deaths by cause, age and population
Variant 3 ag	Proportion disabled by age (age group) and population	-
Variant 3 as	Proportion disabled by age (single-year) and population	-
Variant 4 ag	Proportion disabled by age (age group) and population	Number of deaths by cause, age and population
Variant 4 as	Proportion disabled by age (single-year) and population Proportion disabled by age	Number of deaths by cause, age and population
Variant 5	Attributions by age and population	-

A point of special attention is the consistency of age groups if different input data sets in one analysis (job) are used. The minimum requirements for each variant are specified below. Variants using more data also provide the results of the less data demanding variants. The name of each job is given in brackets.

All input data should be provided by age and population. Age should be included in the input file as a number (e.g. 70) and not a label, such as “70-74” or “70+”.

Variant 1: Only life table data for decomposition of total life expectancy

- number of deaths and population at risk by single of age and population (1-DPs)
- OR
- number of deaths and population at risk by five year age groups and population (1-DPa)
- OR
- life table by single year of age (1-LTs)
- OR
- life table by age group(1-LTa)

It is noteworthy that the output of this variant becomes also available when variants 2-6 are analysed.

Variant 2: Life table and mortality data by cause for decomposition of life expectancy by age and cause of death

- number of deaths and population at risk by single of age and population (2-DPs)

OR

- number of deaths and population at risk by five year age groups and population (2-DPa)

OR

- life table by single year of age (2-LTs)

OR

- abridged life table (2-LTa)

AND

- number of deaths by cause of death and age (age group) and population

The age classes in the cause of death file determine the age classes in the abbreviated life tables that will be constructed if the tool starts from number of deaths and population by single year of age, or from a single year life table.

It is noteworthy that the output of this variant becomes also available when variants 4 and 6 are analysed.

Variant 3: Life table and disability data by age for decomposition of health expectancy by type of effect

- number of deaths and population at risk by single of age and population (3-DPs-as)

OR

- number of deaths and population at risk by five year age groups and population (3-DPa-ag)

OR

- life table by single year of age (3-LTs-as; 3-LTs-ag)

OR

- abridged life table (3-LTa-ag)

AND

- prevalence of disability by single year of age (3-PDs-as, 3-LTs-as) or age group (3DPa-ag, 3-LTa-ag, 3-LTs-ag) and population

If prevalences by single year of age are available, but the life table is abridged (LTa), no information on the population sizes by single year of age is available. Therefore prevalences cannot be aggregated into age groups. For this reason the combination of abridged life table (LTa) or deaths and population by age groups (DPa) and prevalence of disabled by single year of age (as) is not possible.

The user has to specify the age groups in the abbreviated life table for variants starting from single year life table data. These should be the same as the ones used in the file with data on disability prevalence by age group. Also the age groups in the abridged life tables have to match with the prevalence data.

It is noteworthy that the output of this variant becomes also available when variant 6 is analysed.

Variant 4: Life table and disability data by age and mortality data by cause for decomposition of mortality effect by cause

- number of deaths and population at risk by single year of age and population (4-DPa-as, 4-DPa-ag)

OR

- number of deaths and population at risk by five year of age and population (4-PDa-ag)

OR

- life table by single year of age (4-LTs-as; 4-LTs-ag)

OR

- abridged life table (4-LTa-ag)

AND

- prevalence of disability by age (single year (“as” or age group “ag”) and population

AND

- number of deaths by cause of death, age and population

The age classes in the cause of death file determine the age classes in the abbreviated life tables that will be constructed if the tool starts from the number of deaths and population by single year of age, or from a single year life table. The age classes in the file with disability prevalence by age group should be identical to these age groups. Also the age groups in the abridged life tables have to match with the prevalence data.

It is noteworthy that the output of this variant becomes also available when variant 6 is analysed.

Variant 5: Life table data and disability prevalence by cause for decomposition of disability effect by cause

- number of deaths and population at risk by single year of age and population (5-PDs-cause)

OR

- number of deaths and population at risk by five year of age and population (5-PDa-cause)

OR

- life table by single year of age (5-LTs-cause)

OR

- abridged life table (5-LTa-cause)

AND

- prevalence of disability by cause (and age group) and population

The user has to specify the age groups in the abbreviated life table for variants starting from single year life table or deaths and population data (5-DPs-cause; 5-LTs-cause). These should be the same age groups as used in the prevalence by cause data.

It is noteworthy that the output of this variant becomes also available when variant 6 is analysed.

Variant 6: All input data for full decomposition of health expectancy

- number of deaths and population at risk by single year of age and population (6-DPs-cause)

OR

- number of deaths and population at risk by five year of age and population (6-DPa-cause)

OR

- life table by single year of age (6-LTs-cause)

OR

- abridged life table (6-LTa- cause)

AND

- prevalence of disability by cause (and age group) and population

AND

- number of deaths by cause of death and age (age group) and population

The age classes in the cause of death file determine the age classes in the abbreviated life tables that will be constructed if the tool starts from the number of deaths and population by single year of age, or from a single year life table. The age classes in the disability prevalence by cause, should be identical to these age groups. Sequences of causes of death and causes of disability should be the same (not necessarily the labels).

The tool is delivered with example datasets. The dataset with the life table information preferably should contain the names of the populations. These names are used in the output file. The datasets can include more information than actually used in the decomposition analyses, as the user should specify which columns are used. This allows one dataset to be used for more analyses. The only exception is the file with disability data by cause (attribution in mrc file), where all information is used.

A dataset can have more rows before the real data start, but if more than one “title row” is used, a “#” should be given to all except one title line. The program expects the data to start at the second row without #.

We recommend un-experienced users to follow the format of the examples.

3.1.2 The Excel input specification file

The Excel input file is in a csv (comma separate variables) format, that can be saved in excel, choosing “save as”, “csv comma delimited”. If questions pop up your screen (starting with “input file already exists” and “input file may contain features that are not compatible with csv format”, the user answers two times “yes”). As alternative an input specification file in txt format can be used. It should be organised in two columns, separated by a @.

An example of the txt input file is given below:

```
V:\PRJCT\usshare\caspar\decomp12\  @ Q1: Enter the name of your default directory; this will be pasted before all filenames you enter. End always
decomp-1-DEa-boots-txt file.csv @ Q2: Enter the name for output file. You may end this file with ".out"
1 @ Q3: Enter type of input: 1 for population and number of deaths; 2 for single year life table; 3 for abridged
c @ Q3a: enter "t" if this input is given in txt files, "c" if this input is given in csv-files. Use same format
NL-DE-F-2004-5yr.csv @ Q4-1: Enter the name of the files with data specified in Q3 for the FIRST population
IT-DE-F-2004-5yr.csv @ Q4-2: Enter the name of the files with data specified in Q3 for the SECOND population
3.1.5.4 @ Q5-1: Enter column numbers depending on file type, specified in Q3 for file type =1 in Q3 for the FIRST popu
3.1.5.4 @ Q5-1: Enter column numbers depending on file type, specified in Q3 for file type =1 in Q3 for the SECOND pop
NA @ Q6-1: Enter the file name with cause-specific deaths for the FIRST population (if no causes of death, enter
NA @ Q6-2: Enter the file name with cause-specific deaths for the SECOND population (if no causes of death, enter
NA @ Q7: Enter the numbers of the columns with deaths per cause in the files specified in Q8 and Q9; if no cause:
0.1.5.10.15.20.25.30.35.40.45.50.55 @ Q8: If no files with causes of death are included in Q6-1 and Q6-2, enter cut-levels fi
NA @ Q9: enter CAUSE if you have disability data by cause (attributions), AGE if you only have disability data b
NA @ Q10: if you have prevalences of disability by age only (Q9=AGE): enter "S" for single-year prevalences or "G
NA @ Q11-1: enter the filename with disability by cause (arc file from attribution tool) or disability by age (pr
NA @ Q11-2: enter the filename with disability by cause (arc file from attribution tool) or disability by age (pr
NA @ Q12-1: If you have proportion of disability by age: enter column numbers for age (single years or age group)
NA @ Q12-2: If you have proportion of disability by age: enter column numbers for age (single years or age group)
0 @ Q13: Enter age to start life table. For partial life expectancies also enter after ".", the last age group to
0 @ Q14: Enter 0, 1 or 2 to specify the level of detail in the output; 0 for no extra output; 1 for extra output
F @ Q15: Enter T for results in E-notation (for instance 7.048011e-02) or F for rounding to 4 decimals (0.0705)

. Q5-1 and Q5-2
. Enter for file type =1 in Q3 the column numbers for: population, age in single years, mid-year population, deaths, and if applicable ax
. Enter for file type =2 or 3 in Q3 the numbers of the columns with the: population, age, lx, dx, and LX
```

The user may only change the content of the first column. Except for the obligatory questions, the first column may be left empty or NA (not applicable) may be entered. The second column, and more importantly, the rows should never be changed.

The input file the user specifies:

Q1: Enter the name of your default directory. End always with a \

Q2: Enter the name for output file. You may end this file with ".out"

Q3: Enter type of input: 1 for population and number of deaths; 2 for single year life table; 3 for abridged life table

Q3a: enter "t" if this input is given in txt files, "c" if this input is given in csv-files. Use same format for both populations

Q4-1: Enter the name of the files with data specified in Q3 for the FIRST population

Q4-2: Enter the name of the files with data specified in Q3 for the SECOND population

Q5-1: Enter column numbers depending on file type, specified in Q3 for file type =1 in Q3 for the FIRST population (see also row 23 to row 25)

Q5-2: Enter column numbers depending on file type, specified in Q3 for file type =1 in Q3 for the SECOND population (see also row 23 to row 25)

Enter for file type =1 in Q3 the column numbers for: population, age in single years, mid-year population, deaths, and if applicable ax

Enter for file type =2 or 3 in Q3 the numbers of the columns with the: population, age, lx, dx, and LX.

Where:

ax is the fraction of the year alive. If no ax is provided it is set to 0.5.

lx is the number alive at the beginning of the age interval

dx is the number of deaths in the age interval

Lx is the number of person years lived within the age interval

Q6-1: Enter the file name with cause-specific deaths for the FIRST population (if no causes of death, enter "NA")

Q6-2: Enter the file name with cause-specific deaths for the SECOND population (if no causes of death, enter "NA")

Q7: Enter the numbers of the columns with deaths per cause in the files specified in Q8 and Q9

Q8: If no files with causes of death are included in Q6-1 and Q6-2: enter cut-levels for first year of age classes for the abridged life tables. These should match with prevalence age classes if available.

Q9: enter CAUSE if you have disability data by cause (attributions), AGE if you only have disability data by age or NA if you have neither

Q10: if you have prevalences of disability by age only (Q9=AGE): enter "S" for single-year prevalences or "G" for prevalence by age group. If you have no prevalences, enter "NA"

Q11-1: enter the filename with disability by cause (mrc file from attribution tool) or disability by age (proportion) for the FIRST population. If this is a text file, use ".txt"

Q11-2: enter the filename with disability by cause (mrc file from attribution tool) or disability by age (proportion) for the SECOND population. If this is a text file, use ".txt"

Q12-1: If you have proportion of disability by age: enter column numbers for age (single years or age group) and for proportion of disability for the FIRST population IF BOOTSTRAP ALSO NUMBERS OF RESPONDENTS (UNWEIGHTED)".

Q12-2: If you have proportion of disability by age: enter column numbers for age (single years or age group) and for proportion of disability for the SECOND population. IF BOOTSTRAP ALSO NUMBERS OF RESPONDENTS (UNWEIGHTED).

Q13: Enter age to start life table. For partial life expectancies also enter after "," the last age group to be included (age group specified in files with cause-specific deaths or Q8), e.g. "50,70"

Q14: Enter 0, 1, or 2 to specify the level of detail in the output. IF BOOTSTRAP GIVE NUMBER OF REPLICAS

Q15: Enter T for results in E-notation (for instance 7.048011e-02) or F for rounding to 4 decimals (0.0705)

For partial life expectancies you have to enter in Q-13 both the lowest age of the starting age group and last age group. Please be aware that the age groups here are the ones used in the cause-of death files, or, if these files are not used, those specified in Q-8. And be aware that since you work here with age groups, “0,70” means from “0-4” up to and including “70-74”.

3.1.3 Running R

The R-syntax “decomposition” file is ready to use. Users only need to specify the path and name of the input specification file. First open it. This can be done in two ways:

1. In Tinn-R: file / open /, (or double click on name)
2. In R: file/ open script/ open

Then the path and name of the input specification file should be entered or adapted in the line that starts with “filein”.

E.g. `filein <- "d:\\decomman\\decom 1 DP.csv"`

Be careful with the double “\\”

The user can include more input files for different runs. Always the last one will be used by the program. The user can also put “#” before the lines with input files not to be used in the specific run. The program then reads this line as comment.

When running several jobs, we suggest clearing the R console (in Tinn-R: controlling R/clear Console).

3.2 Output of the model

The output depends on the variant of the decomposition tool used and on the selection of optional output on the life table or on the Arriaga method.

We describe here the output of the full decomposition, i.e. both by cause of death and by cause of disability, referred to as “variant 6”. This variant includes also the results of all decompositions that need fewer data. Other variants that provide the described output are given between brackets.

The output of variant 1-5 consists of parts of the output on variant 6.

Part I of the output serves for data inspection and documentation so the user can easily see which data files, population, variables and specifications were used. The user should check this in case an error occurs which is recognized by a blue error message in the console window after running R, but also to assess whether the analysis is correctly specified. It provides:

- Path and name of output file and of input specification file (csv file).
- Path and names of datasets used to construct the life tables (i.e. file with death and population (DPs, DPa), or file with life tables (LTs or LTa), and first rows of these files. In the output, population is here referred to as t and t+n, as in the original publication of Nusselder and Looman (Nusselder and Looman 2004), but population can refer to any type of population, as long as there are two populations specified. We recommend the user to put the name of the population in the file with the life table data, by including a column with in the column heading “population” and in the rows (at least in the first row with data) the name of the population (e.g. Italian males).
- Path and names of datasets with cause-specific deaths for each population (variant 2, 4, and 6 only), and names of the selected causes of death.
- Specification of additional output.
- Specification of age where the life table calculations have to start, and if partial life expectancies, where the life table calculation ends.
- Table with disability by cause (attributions) in each population (variant 5 and 6). As variant 6 includes both causes of death and disability it is important to check that the causes are the same and in the same order. For this reason both are listed.

Part II of the output gives the following results of combining input data and descriptive calculations.

- Listing of the combined information from the different input files by age and population. This includes:
 - deaths by causes and total deaths
 - life table information: l_x , L_x , T_x , dx and mx
 - proportion disabled (π)
- Descriptive calculations of life expectancy, total (all variants), with and without disability (variant 3-6) and differences.

Part III gives supplementary output for the decomposition of life expectancy by cause. It gives:

- Proportion disability by cause
- Life expectancy with disability by cause

In contrast with the other decompositions, which are decompositions of *differences* or *changes* and based on a comparison of two populations, this is the number of years with disability within each population. This output could be easily generated with the tool and is useful as such, but is not part of the decomposition tool that focuses on differences or changes in health expectancy.

Part IV of the output presents the outcomes of the decomposition analysis, starting with:

- Total life expectancy (also variants 1-5) and life expectancy without and with disability (also variant 3-5) in each population
- Differences in total life expectancy (also variants 1-5) and life expectancy without and with disability (also variant 3-5) between the two populations. These differences are the values for the second population minus those for the first population (reference population). Thus, if life expectancy in the first (reference) population is 80 years and in the second population 82, the difference is +2. If the user prefers the other population to be the reference, the user should in the input change the information provided for the first and second population.

Next are the core outcomes from the decomposition analyses given:

- First the decomposition of total life expectancy by age is given (also variant 1-5). This shows which age groups contribute most to the difference in life expectancy. The sum of these age specific contributions is the total life expectancy.
- Next is given the contribution of the disability and mortality effect (also variant 3-5). The mortality effect is the difference in health expectancy due to a difference in the number of person years lived, caused by a difference in total mortality up to and including that age. The disability effect is the difference in health expectancy due to a difference in the proportion with disability.

- Then the disability and the mortality effect by age are given (also variant 3-5). The disability effect is provided by age. It is important to realize that while age-specific differences in prevalence of disability originate from differences in incidence, recovery and mortality rates at younger ages, based on cross-sectional prevalence data, these underlying dynamics cannot be further assessed, and the disability effects by age point where age-specific differences in prevalence occur, but not where they originate.
- Finally the disability effect decomposed by cause (also variant 5) and mortality effect decomposed by cause (also variant 2 and 4) are presented. As variant 6 yields both the decomposition of the disability effect by cause and of the mortality effect by cause, the sum is also given. This yields the contribution of specific causes, both as cause of death and cause of disability to the difference in years with and without disability. For this application it is crucial that the disease entities of the causes of death and causes of disability are comparable.
- The output ends with giving the same information on decomposition by cause in another sequence. Instead of providing separate tables for the mortality, disability and total effect, separate tables are provided for life expectancy without disability, life expectancy with disability and total life expectancy.
- If bootstrap is chosen, at the end the bootstrap results will be presented. The following information is given:
 - Number of replicas
 - Table: short name of the variable
 - Row: row labels of the original output
 - Column: column labels of the original output
 - Value: value from the analyses without bootstrapping
 - Bmean: mean from the analysis with bootstrapping
 - Blow: lower limit of 95% confidence interval
 - Bhig: higher limit of 95% confidence interval

Part 3: Example

The illustration focuses on differences in life expectancy with and without disability between Dutch and Italian women. As disability measure in the example is used the Global Activity Limitation Index (GALI) based on the question “For the past six months at least, to what extent have you been limited because of a health problem in activities people usually do?” As disabled is considered “severely limited” and “limited, but not severely”.

The example illustrates the full decomposition of differences in life expectancy with and without disability by causes of death and disability and age, using data on population and deaths to construct the life table (decom-6_PD_cause). This variant includes also the results of all variants that need fewer data. Variants starting from life tables (LTs or LTa)) instead of the number of deaths and population (DPs or DPa) have similar output.

Example files are available from the authors on request, see Appendix 1. Similar data files can be easily derived from databases available on the internet, including the EHEMU database (<http://www.ehemu.eu/>), Human Mortality database (www.mortality.org), and WHO database (<http://www.who.int/whosis/mort/download/en/index.html>). Only disability data by cause are not generally available and were be derived with the attribution tool based on the SHARE survey (<http://www.share-project.org>). For more details on disability by cause, see the technical report on the attribution tool (Nusselder and Looman 2012).

1. Input

1.1 Input datasets (ASCII or POR-format)

Table 4 shows the names of the datasets for the full decomposition of health expectancy (by age, mortality vs. disability effect and cause), using the number of deaths and population to construct the life table (decom6_PD_cause). The tool always uses two populations, here Dutch and Italian women.

Table 4 Datasets used in the example of decom6_PD_cause

	Netherlands	Italy
Population and deaths	NL-DP-F-2004.csv	IT-DP-F-2004.csv
Prevalence of disability by cause (and age group) and population	NL-F-attrib.mrc	IT-F-artrib.mrc
Number of deaths by cause, age and population	NL-CoD-F-2004-95.csv	IT-CoD-F-2004-95.csv

The mrc file is output of the attribution tool.

The first rows of each file are given below.

For IT-DP-F-2004.csv:

Age	Year	population-2	Deaths	PY at risk
0	2004	itwomen	979	266739.5
1	2004	itwomen	65	264809.5
2	2004	itwomen	50	262843.5
3	2004	itwomen	32	262754
4	2004	itwomen	31	263026.5
5	2004	itwomen	22	259776.5
6	2004	itwomen	26	258421

Including the population name, here as example itwomen, is recommended as this name is printed in the output file.

For IT-F-attrib.mrc:

	50	55	60	65	70	75	80	85
nn	183.346	243.013	212.2369	186.645	202.062	140.755	142.944	61.9988
disab	51.968	84.031	78.5461	83.909	115.816	81.486	106.523	50.1649
backgrnd	18.597	23.309	22.3606	32.236	40.080	30.813	50.653	33.5817
heart	1.839	5.660	5.5544	5.306	9.039	6.905	12.065	2.7869
stroke	1.070	1.147	0.6353	1.341	4.242	2.302	2.089	0.0000
cancer	1.708	4.637	3.3836	4.447	3.809	2.169	1.860	1.5533
copd	1.753	2.475	1.4183	1.636	2.153	1.414	1.292	0.5957
diabetes	2.384	5.209	7.9318	7.710	13.142	7.856	8.623	1.0075
muskulo	15.781	28.789	25.6613	25.614	31.451	21.961	18.605	7.4829
other	8.835	12.805	11.6009	5.620	11.900	8.067	11.336	3.1570

The mrc file is made with the attribution tool and can be opened in Notepad or TextPad.

For IT-CoD-2004-95.csv:

Age	Heart	Stroke	cancer	COPD	diabetes	muskul	other	All
0	24	5	6	0	0	1	952	988
1	10	3	34	0	0	2	132	181
5	4	2	39	0	1	0	72	118

1.2 Input specification file (.csv)

The input specification file of variant 6 (input decomposition 6-DPA-cause.csv) is given in Table 1. Appendix 2 shows the names of the input specification files for the examples illustrating the other variants of the decomposition tool. The zip-file including several examples of cvs and output files is available from the authors.

Table 1: Example of input specification file

Q7: excel changes “2:8” into “02:08”.

1.3 R syntax file (.R)

The R program specifies the location (path) and name of the input specification file. Different input specification files can be included, for example:

```
filein <- "D:\\decomphandleiding\\input decomposition 1-DPs.csv"
filein <- "D:\\decomphandleiding\\input decomposition 2-DPs.csv"
filein <- "D:\\decomphandleiding\\input decomposition 3-DPs-ag.csv"
filein <- "D:\\decomphandleiding\\input decomposition 4-DPs-ag.csv"
filein <- "D:\\decomphandleiding\\input decomposition 6-DPs-cause.csv"
```

Here the R program uses the last one, thus “input decomposition 6-DPs-cause.csv” and creates the output file specified in Q2 in each input specification file. The double “\\” and quotes “” are necessary.

2. Output

Part I of the output serves for data inspection and documentation

Each output file with the location (path) and name of the output file and name of the csv file, for example.

```
[1] "> > > OUTPUT PART 1: DATA INSPECTION AND DOCUMENTATION < < <"
[1] "d:\\decomman\\decomp-6-DP-cause.out"
[1] "D:\\decomman\\input decomposition 6-DP-cause.csv"
```

This is followed by the input files that are used to make the life tables and the first rows of these files

```
[1] "files with deaths and years-at-risk:"
[1] "d:\\decomman\\NL-DP-F-2004.csv"
[1] "d:\\decomman\\IT-DP-F-2004.csv"

[1] "label, age, Px and Dx for population t:"
  population.1 age PY.at.risk deaths
1      nlwomen  0   96178.5    362

[1] "label, age, Px and Dx for population t+n:"
  population.2 age PY.at.risk deaths
1       itwomen  0  266739.5    979
```

This shows whether the right input is specified. Population is referred to as t and t+n, as in the original publication of Nusselder and Looman (Nusselder and Looman 2004), but can refer to any population.

We recommend the user include the name of the population in the file with the life table data, e.g. nlwomen in the file with the life table data.

Next are presented the names of the input files with information on causes of death, and the selected causes of death. The age intervals in the cause of death data files determine the age intervals in the decomposition and are the same as for the attributions (mrc file).

```
[1] "filename for cause specific deaths for population t:"
[1] "d:\\decomman\\NL-CoD-F-2004-95.csv"

[1] "filename for cause specific deaths for population t+n:"
[1] "d:\\decomman\\IT-CoD-F-2004-95.csv"

[1] "selected causes for cause specific deaths:"
[1] "Heart" "stroke" "cancer" "COPD" "diabetes" "muskul" "other"
```

Then the user choice regarding the additional output and starting age of the life table is printed. If problems arise or results are unclear, using “1” in Q-14 in the input specification file gives additional output on the life table calculations and using “2” gives additional output on the Arriaga analyses. Here we selected no additional output.

```
[1] "no extra output"

[1] "age at which to start the life table:"
[1] 50
```

It is noteworthy that in this example age 50 is the lowest possible age as data on disability by cause (attributions) start at age 50.

For partial life expectancies also the age to stop the life table is given (not included here).

This is followed by a print of the attribution table including information on disability by cause.

```
[1] "attributions for populations (t) and (t+1):"
      50      55      60      65      70      75      80      85
nn    305.000 350.000 251.000 189.000 173.000 115.000 69.0000 59.0000
disab 132.704 160.721 131.877 90.496 85.240 74.312 45.2233 37.0250
backgrnd 74.425 69.476 58.865 28.644 29.681 28.016 16.5044 11.3208
HEART  2.121  5.763  3.382  4.034  2.919  5.192  2.0211  2.0732
STROKE 1.240  2.834  1.955  2.295  3.059  2.641  1.2043  2.5925
CANCER  2.363  3.106  3.501  2.264  1.816  1.549  0.9695  0.8185
COPD   10.072 15.454  7.852  7.432  3.768  5.098  2.6250  1.8959
DIABETES 3.523  3.834  5.156  4.491  4.584  4.251  1.7456  1.1511
MUSKULO 17.454 31.035 27.057 21.406 22.254 15.893 10.9917 9.9782
OTHER  21.507 29.219 24.108 19.931 17.160 11.672  9.1618  7.1947
      50      55      60      65      70      75      80      85
nn    183.346 243.013 212.2369 186.645 202.062 140.755 142.944 61.9988
disab  51.968 84.031 78.5461 83.909 115.816 81.486 106.523 50.1649
backgrnd 18.597 23.309 22.3606 32.236 40.080 30.813 50.653 33.5817
heart   1.839  5.660  5.5544  5.306  9.039  6.905 12.065  2.7869
stroke  1.070  1.147  0.6353  1.341  4.242  2.302  2.089  0.0000
cancer  1.708  4.637  3.3836  4.447  3.809  2.169  1.860  1.5533
copd    1.753  2.475  1.4183  1.636  2.153  1.414  1.292  0.5957
diabetes 2.384  5.209  7.9318  7.710 13.142  7.856  8.623  1.0075
muskulo 15.781 28.789 25.6613 25.614 31.451 21.961 18.605  7.4829
other   8.835 12.805 11.6009  5.620 11.900  8.067 11.336  3.1570
```

As variant 6 includes both causes of death and disability it is important to check that the causes are the same and in the same order. For this reason both are listed in the output.

```
[1] "selected causes for disability attributions (next to background):"
[1] "HEART" "STROKE" "CANCER" "COPD" "DIABETES" "MUSKULO" "OTHER"
[1] " ... should match with:"
[1] " causes of death:"
[1] "Heart" "stroke" "cancer" "COPD" "diabetes" "muskul" "other"
```

Part II of the output gives results of combining input data and descriptive calculations.

First the combined information from the different input files for each population it presented. This includes:

- deaths in the life table population, by cause and total
- life table parameters: including: l_x , L_x , T_x , dx , mx and ex
- proportion disabled (π)

```

[1] "-----"
[1] ">> OUTPUT PART 2: COMBINING INPUT DATA AND DESCRIPTIVE CALCULATIONS <<<"
[1] "nlwomen" "itwomen"

[1] "lifetables starting at 50 (t) and (t+1)"
ageclass Heart stroke cancer COPD diabetes muskul other All lx Lx Tx ex Px Dx dx ax
50 50 209 67 1081 41 31 5 387 1821 1.00000000 4 9610612 3193466 17 32 949635 550578 0 1821 0 01642221 0 003307433 0 43509
55 55 296 94 1539 51 47 10 471 2508 0.98357779 4 9657369 2712641 39 28 455802 543176 0 2508 0 02267861 0 004617288 0 45320
60 60 387 110 1557 114 54 19 535 2776 0.9608919 4 9265005 2241056 61 24 063749 403946 0 2776 0 03261994 0 006872206 0 62540
65 65 452 193 1868 171 110 25 726 3745 0.92827925 4 5278266 1782965 97 19 817675 345761 0 3745 0 04923565 0 01081181 0 47881
70 70 1313 490 2238 303 217 50 1238 5849 0.87904368 4 2126811 1344130 74 15 775817 314011 0 5849 0 07841575 0 018626736 0 43271
75 75 2200 908 2658 383 362 90 2370 8971 0.80062785 3 7174419 935839 28 12 060321 269855 0 8971 0 12362961 0 033243779 0 64619
80 80 3649 1558 2861 484 515 137 4359 13567 0.67698624 2 9440502 575546 22 8 771843 213467 0 13567 0 18809362 0 06355491 0 65541
85 85 3645 1636 1975 416 432 107 5137 13348 0.48990462 1 8917456 290209 38 6 124583 113734 5 13348 0 22225642 0 117361047 0 62754
90 90 2853 1061 1018 245 297 75 4399 9948 0.2664820 0 8508025 106860 78 4 134928 48065 0 9948 0 17750782 0 206969728 0 62754
95 95 1034 406 257 91 89 29 2144 4050 0.08914038 0 2517686 24401 32 2 824406 11630 0 4050 0 09533186 0 348237317 0 62754

ageclass Heart stroke cancer COPD diabetes muskul other All lx Lx Tx ex Px Dx dx ax
50 50 442 185 2751 95 187 60 782 4442 1.0000000 4 9737617 3417903 89 35 019047 1884369 5 4189 0 01104504 0 0022223025 0 28344
55 55 787 273 3848 76 310 74 1077 6445 0.9889550 4 9063325 2932458 37 30 380843 1885207 5 6249 0 01639498 0 003314754 0 34578
60 60 1322 487 5509 130 687 159 1392 9696 0.9725600 4 8046423 2453594 02 25 848228 1732403 0 8946 0 02467318 0 005183925 0 37008
65 65 2668 974 7206 275 1284 226 2282 14915 0.9478868 4 6511245 1984654 77 21 452256 1711431 0 14032 0 03827874 0 008198987 0 44956
70 70 5546 2345 9530 627 2125 398 4052 24623 0.9096081 4 4039563 1530699 05 17 241696 1890951 5 22500 0 06239387 0 014192765 0 57317
75 75 11013 5089 11999 1160 2513 680 8314 40788 0.8472145 3 9956447 1100887 26 13 313101 1397124 0 35708 0 10515648 0 025273874 0 57892
80 80 17697 8627 11759 1727 2409 690 12444 53535 0.7420580 3 3096786 710887 21 9 815385 1078260 5 55649 0 17355085 0 051609977 0 74520
85 85 21544 10317 7919 2039 1779 971 13656 57825 0.5685072 2 3108522 387858 34 6 990072 464340 5 45825 0 22162993 0 098680355 0 80912
90 90 22100 3515 5098 1873 520 461 14375 53942 0.2468772 1 1898955 162345 48 4 775024 285310 5 48854 0 20617541 0 171230992 0 80912
95 95 9103 3453 1191 782 114 183 6432 21258 0.1407008 0 4734785 46212 11 3 365144 69640 0 19869 0 15074425 0 285310167 0 80912

[1] "table with attributions is extended by 2 age classes, because of the life table"

```

Next is given the differences in cause-specific disability prevalence between two populations, i.e., Dutch females and Italian females. In this example Italian females have higher disability due to musculoskeletal diseases at age 50-54 (NL: 0.05722580; IT: 0.08607389, difference 0.02884809). The original data can be derived from the input file with cause-specific disability data (15.781/ 183.346 =0.08607), and is also presented in the next part with supplementary data.

```

[1] "Cause-specific difference in proportion disabled:"
backgrnd heart stroke cancer copd diabetes muskulo other
50 -0.14258143 0.003076363 0.001770530 0.001568417 -0.02346136 0.001452256 0.02884809 -0.02232572
55 -0.10258618 0.006825221 -0.003377231 0.010206998 -0.03396965 0.010480782 0.02979548 -0.03079020
60 -0.12916703 0.012696516 -0.004795554 0.001994229 -0.02460049 0.016830353 0.01311097 -0.04138846
65 0.02115697 0.007084281 -0.004958046 0.011846992 -0.03055712 0.017546232 0.02397414 -0.07534358
70 0.02679050 0.027961160 -0.003311683 0.008353663 -0.01112495 0.038542644 0.02701592 -0.04029677
75 -0.02470778 0.003908444 -0.006610759 -0.001940000 -0.03428473 0.019847390 -0.01782096 -0.04418400
80 0.11516188 0.055112443 -0.002839470 -0.001038605 -0.02900489 0.035025828 -0.02914378 0.05347563
85 0.34977124 0.009811702 -0.043940797 0.011180742 -0.02252575 -0.003259940 -0.04842846 -0.07102416
85 0.34977124 0.009811702 -0.043940797 0.011180742 -0.02252575 -0.003259940 -0.04842846 -0.07102416
85 0.34977124 0.009811702 -0.043940797 0.011180742 -0.02252575 -0.003259940 -0.04842846 -0.07102416

```

The last output of part II give the results of the calculations of life expectancy, total (all variants), with and without disability (variant 3-6) and differences in life and health expectancies. This is the starting point of the decomposition analyses.

```

[1] " recapitulation of overall life expectancies"

[1] "LE, DFLE and years with disability from 50 : "
without with total
nlwomen 15.61895 17.33069 32.94963
itwomen 17.52443 17.49462 35.01905

[1] " difference in LE:"
without with total
1.9054845 0.1639272 2.0694117

```

Part III gives supplementary output for the decomposition of life expectancy by cause. It gives the proportion disability by cause for each population. Here we show this output for Dutch women.

```
[1] "proportion disability by cause and years with disability by cause:"

[1] "nlwomen"
[1] "proportion with disability by cause:"
  backgrnd      HEART      STROKE      CANCER      COPD      DIABETES
50 0.2440146 0.006954046 0.004065543 0.007747483 0.03302270 0.01155073
55 0.1985029 0.016465714 0.008097143 0.008874286 0.04415429 0.01095429
60 0.2345237 0.013474206 0.007788904 0.013948313 0.03128311 0.02054199
65 0.1515539 0.021343679 0.012142723 0.011978704 0.03932232 0.02376164
70 0.1715645 0.016872634 0.017681873 0.010496987 0.02178009 0.02649680
75 0.2436174 0.045147826 0.022965217 0.013469565 0.04433043 0.03696522
80 0.2391937 0.029291240 0.017453585 0.014050694 0.03804339 0.02529849
85 0.1918785 0.035139078 0.043940797 0.013872919 0.03213399 0.01951022
85 0.1918785 0.035139078 0.043940797 0.013872919 0.03213399 0.01951022
85 0.1918785 0.035139078 0.043940797 0.013872919 0.03213399 0.01951022
  MUSKULO      OTHER
50 0.05722580 0.070514222
55 0.08867143 0.08348286
60 0.10779763 0.09604854
65 0.11325801 0.10545386
70 0.12863433 0.09918959
75 0.13820000 0.10149565
80 0.15929965 0.13277942
85 0.16912249 0.12194440
85 0.16912249 0.12194440
85 0.16912249 0.12194440
```

Next the life expectancy with disability by cause is presented. In contrast with the other decompositions, which are decompositions of differences or changes and based on a comparison of two populations, this is the number of years with disability by cause within each population. The output for Dutch females shows that of the 17.3 years Dutch women live with disability, 3.78 are due to musculoskeletal disorders.

```
[1] "LE with disability by cause:"
  backgrnd      HEART      STROKE      CANCER      COPD      DIABETES      MUSKULO
[1,] 6.878247 0.7053102 0.4941801 0.3789788 1.169345 0.6972237 3.768712
      OTHER
[1,] 3.238691
[1] "LE with disability:"
[1] 17.33069
```

This output can be easily generated with the tool and is presented for this reason. It does not result from the decomposition tool that decomposes differences or changes in health expectancy.

Part IV presents the outcomes of the decomposition analysis.

First is given again the starting point of the decomposition analysis.

```
"=====
[1] "> > > OUTPUT PART 4: OUTCOMES OF DECOMPOSITION ANALYSIS < < <"

[1] "LE, DFLE and years with disability from 50 : "
      without      with      total      %HLY/LE
nlwomen 15.61895 17.33069 32.94963 47.40249
itwomen 17.52443 17.49462 35.01905 50.04257
```

This shows that Dutch women have a total life expectancy at age 50 of 32.9 years, of which 15.6 are lived free of disability (47.4%) and 17.3 with disability. Italian women live longer; their total life expectancy at age 50 is 35.0 years, of which 17.52 without disability (50%) and 17.49 years with disability.

The differences between the two populations are also shown:

```
[1] " difference in LE:"
      without      with      total
1.9054845 0.1639272 2.0694117
```

These differences are the values for the second population (Italian females) minus those for the first population (reference population, Dutch females).

Before the decompositions of the health expectancies are presented (only variant 3-6), the decomposition of the difference in total life expectancy by age, is given (each variant).

```
[1] " LE difference decomposed by age of origin:"
      total
y=50 0.17088798
y=55 0.17341481
y=60 0.19173631
y=65 0.22329324
y=70 0.27771492
y=75 0.29180051
y=80 0.28297438
y=85 0.25503548
y=90 0.14041226
y=95 0.06214185
TOTAL 2.06941173
```

This shows how much each age group contributed to the 2.069 difference in life expectancy between Dutch and Italian women. For instance, age group 75-79 contributed 0.2918 years to this difference. The sum of these age specific contributions is the difference in total life expectancy.

Next is given the decomposition of differences in years with and without disability by type of effect, i.e. mortality effect and disability effect.

```
[1] " HE difference decomposed into mortality and disability effect:"
      without      with      total
Mort 0.7242226 1.3451891 2.069412
Disab 1.1812619 -1.1812619 0.000000
TOTAL 1.9054845 0.1639272 2.069412
```

“Mort” is mortality effect and is the difference in years without and years with disability that is due to differential mortality. Thus of the 1.9 longer life expectancy without disability of Italian women, 0.72 is due to differences in mortality. This means that lower mortality in Italian women results in a 0.72 years longer disability-free life expectancy in Italian women as compared to Dutch women, and in a 1.34 years longer life expectancy with disability.

“Disab” is disability effect and is the difference in years without and with disability that is due to differences in the prevalence of disability. Of the 1.9 years difference in years without disability, 1.18 years is due to differences in disability prevalence. Thus lower disability in Italian women explains 1.18 years out of the 1.9 years longer disability free life expectancy in Italian women. The difference in the number of years with disability is only 0.163 years. This is the net result of an opposite mortality (+1.35) and disability effect (-1.18). Lower mortality of Italian women increases the number of years with disability by 1.35 years as compared to Dutch females, while the lower disability prevalence reduces the number of years with disability by 1.18 years.

Differences in total life expectancy are by definition only caused by differences in mortality.

What follows is the disability and mortality effect by age. This refers to age-groups where differences in disability prevalence occur. The decomposition of the mortality effect by age shows the age groups where the mortality differences originate.

```
[1] "disability effect by age:"
      without      with
50-54  0.75332208 -0.75332208
55-59  0.55414851 -0.55414851
60-64  0.74018596 -0.74018596
65-69  0.13424277 -0.13424277
70-74 -0.34662082  0.34662082
75-79  0.25943153 -0.25943153
80-84 -0.28078545  0.28078545
85-89 -0.38153805  0.38153805
90-94 -0.18527782  0.18527782
95 +   -0.06584685  0.06584685
TOTAL  1.18126188 -1.18126188

[1] " mortality effect by age:"
      without      with      total
y=50  0.08138208  0.08950590  0.17088798
y=55  0.07783327  0.09558154  0.17341481
y=60  0.08106788  0.11066843  0.19173631
y=65  0.08661306  0.13668018  0.22329324
y=70  0.09677158  0.18094334  0.27771492
y=75  0.09053812  0.20126238  0.29180051
y=80  0.08112952  0.20184486  0.28297438
y=85  0.07183465  0.18320082  0.25503548
y=90  0.03954927  0.10086300  0.14041226
y=95  0.01750320  0.04463864  0.06214185
TOTAL  0.72422263  1.34518910  2.06941173
```

The disability effects by age for life expectancy without and with disability have by definition opposite signs. Disability effect by age shows positive signs for the younger ages and negative signs for the older ages. At younger more favorable disability prevalence in Italian women increase life expectancy without disability, while at older ages the opposite is true.

The mortality effect by age indicates that at all ages, lower mortality rates in Italian women increase life expectancy with and without disability of Italian women as compared to Dutch women (*ceteris paribus*). It also shows which age groups contribute most, For instance, a large part of the differences

in life expectancy with disability that are due to difference in mortality, originate between ages 70 and 89 years of age.

Next is shown the disability effect and mortality effect decomposed by cause.

```
[1] " disability effect by cause:"
      without      with total
backgrnd 0.1290791 -0.1290791  0
heart    -0.4832713  0.4832713  0
stroke   0.2265153 -0.2265153  0
cancer   -0.2007173  0.2007173  0
copd     0.8893228 -0.8893228  0
diabetes -0.5560604  0.5560604  0
muskulo  -0.3866533  0.3866533  0
other    1.5630469 -1.5630469  0
TOTAL    1.1812619 -1.1812619  0
```

This output for instance shows that while for Italian women overall lower disability prevalence as compared to Dutch women increases life expectancy with and without disability (*ceteris paribus*), some diseases, such as heart diseases, diabetes and musculoskeletal yield higher disability prevalence. This can be due to a higher prevalence of these diseases or due to a higher disabling impact among Italian women.

The decomposition of the mortality effect by cause of death shows that mainly other causes of death are responsible for the differences in year with and without disability between Italian and Dutch women, as far as caused by mortality differences.

```
[1] " mortality effect by cause"
      without      with      total
Heart    0.045764427  0.04162691  0.08739134
stroke  -0.018456524 -0.05742821 -0.07588473
cancer   0.287992130  0.47091189  0.75890402
COPD     0.054155086  0.09257280  0.14672789
diabetes -0.035279506 -0.04566264 -0.08094215
muskul  -0.009153184 -0.01424478 -0.02339796
other    0.399200204  0.85741311  1.25661331
TOTAL    0.724222632  1.34518910  2.06941173
```

Since both the mortality effect and disability effect are available by cause, the sum is given. The yields the contribution of specific causes, both as cause of death and cause of disability, to the difference in years with and without disability. It is crucial that the disease entities of the causes of death and causes of disability are comparable.

```
[1] " sum mortality and disability effect by cause:"
      without      with      total
backgrnd 0.12907913 -0.1290791  0.00000000
heart    -0.43750683  0.5248982  0.08739134
stroke   0.20805878 -0.2839435 -0.07588473
cancer   0.08727487  0.6716292  0.75890402
copd     0.94347790 -0.7967500  0.14672789
diabetes -0.59133993  0.5103978 -0.08094215
muskulo  -0.39580653  0.3724086 -0.02339796
other    1.96224712 -0.7056338  1.25661331
TOTAL    1.90548451  0.1639272  2.06941173
```

The last part of the output presents the same information in a different way.

3. Illustration of features of the tool

3.1 Illustration of different starting ages

The decomposition analyses can start at different ages. In the example we started at age 50 as disability data by cause that were used as input started at age 50. This was specified in Question 13 “age to start the life table”. For the purpose of illustration, we show a selection of the outcomes for starting age 50 and starting age 60.

```
[1] "age at which to start the life table:"
[1] 50

[1] "LE, DFLE and years with disability from 50:"
      without   with   total %HLY/LE
nlwomen 15.61895 17.33069 32.94963 47.40249
itwomen 17.52443 17.49462 35.01905 50.04257

[1] " difference in LE:"
      without   with   total
1.9054845 0.1639272 2.0694117

[1] " LE difference decomposed by age of origin:"
      total
y=50 0.17088798
y=55 0.17341481
y=60 0.19173631
y=65 0.22329324
y=70 0.27771492
y=75 0.29180051
y=80 0.28297438
y=85 0.25503548
y=90 0.14041226
y=95 0.06214185
total 2.06941173

[1] "age at which to start the life table:"
[1] 60

[1] "LE, DFLE and years with disability from 60 :":
      without   with   total %HLY/LE
nlwomen 10.59949 13.46426 24.06375 44.04755
itwomen 11.05398 14.79424 25.84823 42.76496

[1] " difference in LE:"
      without   with   total
0.4544931 1.3299863 1.7844793

[1] " LE difference decomposed by age of origin:"
      total
y=60 0.19837728
y=65 0.23102606
y=70 0.28732547
y=75 0.30188366
y=80 0.29272680
y=85 0.26377257
y=90 0.14517162
y=95 0.06419585
total 1.78447931
```

3.2 Illustration of partial life expectancies

Generally only the starting age is quantified in question 13. For partial life expectancies both the starting and stop age should be entered. The stop age is the first age of the last age interval to be included in the analysis. Here we show the outcomes of the partial life expectancy between age 50 and 74 (Q13: 50,70).

```
[1] "age at which to end the life table (partial LE):"  
[1] 70  
  
[1] "LE, DPLE and years with disability between 50 and 70:"  
      without      with      total      %HLY/LE  
nlwomen 12.17393 11.11988 23.29381 52.26252  
itwomen 14.24016  9.49966 23.73982 59.98427  
  
[1] " difference in LE:"  
      without      with      total  
2.0662269 -1.6202157  0.4460111  
  
[1] " LE difference decomposed by age of origin:"  
      total  
y=50 0.11383018  
y=55 0.10426659  
y=60 0.09930635  
y=65 0.08456089  
y=70 0.04404712  
total 0.44601113
```

3.2 Illustration of confidence intervals with bootstrapping

For the variants starting from population and death numbers (DPs, DPa) it is possible to use bootstrapping to obtain confidence intervals.

For bootstrapping in combination with prevalence data by age (variant 3 or 4) the number of respondents in the survey by age and population should be given in question Q12-1 and Q12-2.

For bootstrapping in combination with cause of death data, the causes of death should be specified as numbers, not as proportions.

For bootstrapping in combination with prevalence data by cause (variant 6) the mrcxx obtained from bootstrapping in the attribution tool should be provided, next to the mrc file. The number of bootstraps runs in the attribution tool (to make mrcxx) and in the decomposition tool should be identical.

Appendices

Appendix 1: Data files used in the illustrations of the different variants and options

A zip file with the data files, input specification files and output files is available from the author's on request.

Data files

Life table data	Option: Death and population data	Option: Life table by single year of age	Abridged life table
ALL variants	NL-DP-F-2004.csv IT-DP-F-2004.csv NL-DP-F-2004-5yr.csv IT-DP-F=2004-5yr.csv	NL-LTs-F-2004-95.txt IT-LTs-F-2004-95.txt	NL-LTa-F-2004.csv NL-LTa-F-2004.csv

ADDITIONAL INFORMATION		
Additional information	DISABILITY	CAUSE OF DEATH
Variant 1	-	-
Variant 2	-	NL-CoD-2004-95.csv IT-CoD-2004-95.csv
Variant 3 ag	NL-prevdis-5year-F.csv IT-prevdis-5year-F.csv NL-prevdis-5year-F-boot.csv IT-prevdis-5year-F-boot.csv	-
Variant 3 as	NL-prevdis-1year-F.csv IT-prevdis-1year-F.csv NL-prevdis-1year-F-boot.csv IT-prevdis-1year-F-boot.csv	-
Variant 4 ag	NL-prevdis-5year-F.csv IT-prevdis-5year-F.csv NL-prevdis-5year-F-boot.csv IT-prevdis-5year-F-boot.csv	NL-CoD-2004.csv IT-CoD-2004.csv
Variant 4 as	NL-prevdis-1year-F.csv IT-prevdis-1year-F.csv NL-prevdis-5year-F-boot.csv IT-prevdis-5year-F-boot.csv	NL-CoD-2004.csv IT-CoD-2004.csv

Variant 5	NL-F-attrib.mrc IT-F-attrib.mrc NL-F-attrib.mrcxx IT-F-attrib.mrcxx	-
Variant 6	NL-F-attrib.mrc IT-F-attrib.mrc NL-F-attrib.mrcxx IT-F-attrib.mrcxx	NL-CoD-2004.csv IT-CoD-2004.csv

* For some files we included more variants with different number of age groups, so ensure that all files within one job have the same length. The last part of the file name, such as “-85” refers to the last age group included.

Input specification files

Without bootstrapping

Variant	DPa	DPs	LTs	LTa
1	input decomposition 1-DPa.csv	input decomposition 1-DPs.csv	input decomposition 1-LTs.csv	input decomposition 1-LTa.csv
2	input decomposition 2-DPa.csv	input decomposition 2-DPs.csv	input decomposition 2-LTs.csv	input decomposition 2-LTa.csv
3	input decomposition 3-DPa-ag.csv	input decomposition 3-DPs-as.csv input decomposition 3-DPs-ag.csv	input decomposition 3-LTs as.csv input decomposition 3-LTs ag.csv	input decomposition 3-LTa ag.csv
4	input decomposition 4-DPa-ag.csv	input decomposition 4-DPs-as.csv input decomposition 4-DPs-ag.csv	input decomposition 4-LTs as.csv input decomposition 4-LTs ag.csv	input decomposition 4-LTa ag.csv
5	input decomposition 5-DPa-cause	Input decomposition 5-DPs-cause	input decomposition 5-LTs cause.csv	input decomposition 5-LTcause.csv
6	input decomposition 6-DPa-cause	Input decomposition 6-DPs-cause	input decomposition 6-LTs cause.csv	input decomposition 6-LTcause.csv

Options	DP
1 Different starting age	input decomposition 6-DPa-cause-starting age input decomposition 6-DPs-cause-starting age
2 Partial Life expectancy	input decomposition 6-Dpa-cause-partial input decomposition 6-Dps-cause-partial
3 Extra output life table	input decomposition 6-DPa-cause-output LT input decomposition 6-DPs-cause-output LT
4 Extra output Arriaga	input decomposition 6-DPa-cause-output Ar input decomposition 6-DPs-cause-output Ar

With bootstrapping (only variant DP)

Variant	DPa boot	DPs boot
1	input decomposition 1-DPa-boot.csv	input decomposition 1-DPs-boot.csv
2	input decomposition 2-DPa-boot.csv	input decomposition 2-DPs-boot.csv
3	input decomposition 3-DPa-ag-boot.csv	input decomposition 3-DPs-as-boot.csv input decomposition 3-DPs-ag-boot.csv
4	input decomposition 4-DPa-ag-boot.csv	input decomposition 4-DPs-as-boot.csv input decomposition 4-DPs-ag-boot.csv
5	input decomposition 5-DPa-cause-boot.csv	input decomposition 5-DPs-cause-boot.csv
6	input decomposition 6-DPa-cause.boot.csv	input decomposition 6-DPs-cause-boot.csv

Output files

Variant	DPs	DPa
1	decom- 1-DPs.out	decom- 1-DPa.out
2	decom- 2-DPs.out	decom- 2-DPa.out
3	decom- 3-DPs-ag.out	decom- 3-DPa-as.out decom- 3-DPa-ag.out
4	decom- 4-DPs-ag.out	decom- 4-DPa-as.out decom- 4-DPa-ag.out
5	decom- 5-DPs-cause.out	decom- 5-DPa-cause.out
6	decom- 6-DPs-cause.out	decom- 6-DPa-cause.out

Options	DP
1 Different starting age	Decom- 6-DPs-cause-starting age.csv Decom- 6-DPa-cause-starting age.csv
2 Partial Life expectancy	Decom- 6-DPs-cause-partial.csv Decom- 6-DPa-cause-partial.csv
3 Extra output life table	Decom- 6-Dps-output LT.csv Decom- 6-Dpa-output LT.csv
4 Extra output Arriaga	Decom- 6-DPs-output Ar.csv Decom- 6-DPa-output Ar.csv

Variant	DPs	DPa
1	decom- 1-DPs-boot.out	decom- 1-DPa-boot.out
2	decom- 2-DPs-boot.out	decom- 2-DPa-boot.out
3	decom- 3-DPs-as-boot.out decom- 3-DPs-ag-boot.out	decom- 3-DPa-ag-boot.out
4	decom- 4-DPs-as-boot.out decom- 4-DPs-ag-boot.out	decom- 4-DPa-ag-boot.out
5	decom- 5-DPs-cause-boot.out	decom- 5-DPa-cause-boot.out
6	decom- 6-DPs-cause-boot.out	decom- 6-DPa-cause-boot.out

Appendix 3: Further working with the output

For those who want to continue working in R after running `decomposition.r` we give some information in this paragraph. All results are stored in the list called `RESULT`. This contains four more lists as can be seen after entering `names(RESULT)`.

```
names(RESULT)
```

```
[1] "Outlifetab" "OutArriaga" "summary" "bootstrap"
```

The content of each list can be seen and further used:

```
names(RESULT$Outlifetab)
```

```
[1] "Lftab-nlwomen" "Lftab-itwomen" "Lfpop-nlwomen" "Lfpop-itwomen"
```

The names `nlwomen` and `itwomen` refer to the two populations, examined in variant `decom- 6-DPa-cause`.

```
names(RESULT$OutArriaga)
```

```
[1] "Cmort-nlwomen" "DE-nlwomen" "IE-nlwomen" "OE-nlwomen" "I-nlwomen"
```

```
[6] "TOT-nlwomen" "colsumsTOT-nlwomen" "CTOT-1.vs.2" "Cmort-itwomen" "DE-itwomen"
```

```
[11] "IE-itwomen" "OE-itwomen" "I-itwomen" "TOT-itwomen" "colsumsTOT-itwomen"
```

```
[16] "CTOT-2.vs.1" "CTOT" "Mort" "Dis" "Disage"
```

```
names(RESULT$summary)
```

```
[1] "LE" "DiffLE" "LEorigin" "Byeffect" "Disage" "ByageM" "BycauseD" "BycauseM"
```

```
"Bycause" "Bycause 1"
```

```
[11] "Bycause 2" "Bycause 3"
```

```
names(RESULT$bootstrap)
```

```
[1] "CIboots" "Outboots"
```

Each of these objects contains a heading and a content, like:

```
> RESULT$summary$LE
```

This yield for instance:

```
[1] "LE, DFLE and years with disability from 50"
```

```
[[2]]      without      with      total      %HLY/LE
nlwomen   15.54704   17.40259   32.94963   47.18426
itwomen   17.59378   17.42527   35.01905   50.24060
```

All (extra) output from the program can be found in this way and used as input for plots etc. Take care to use quotes if the name contains spaces or –signs.

References

Andreev, E., V. Shkolnikov, et al. (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.

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

Arriaga, E. E. (1989). Changing trends in mortality decline during the last decades. Differential mortality: methodological issues and biological factors. L. Ruzicka, G. Wunsch and P. Kane. Oxford, Clarendon Press: 105-129.

Barendregt, J. J., L. Bonneux, et al. (1994). "Health expectancy: an indicator for change?" Journal of Epidemiology and Community Health 48(5): 482-487.

Efron, B. and R. J. Tibshirani (1994). An introduction to the bootstrap Palto Alto, California.

Jagger, C. (1999). Health expectancy calculations by the Sullivan method: A practical guide.

Mathers, C. D. and J. M. Robine (1997). "How good is Sullivan's method for monitoring changes in population health expectancies?" J Epidemiol Community Health 51(1): 80-86.

Nusselder, W. and C. W. Looman (2010). Technical Report on Attribution Tool. Rotterdam, Erasmus MC.

Nusselder, W. and C. W. Looman (2012). Technical Report on Attribution Tool, version 2012. Rotterdam, Erasmus MC.

Nusselder, W. J. and C. W. Looman (2004). "Decomposition of differences in health expectancy by cause." Demography 41(2): 315-334.

Pollard, J. H. (1988). "On the decomposition of changes in expectation of life and differentials in life expectancy." Demography 25(2): 265-276.

Pressat, R. (1985). "Contribution des écarts de mortalité par âge à la différence des vies moyennes." Population, 4-5, 766-770(4-5): 766-770.

Shkolnikov, V. M., E. M. Andreev, et al. (2006). "The changing relation between education and life expectancy in central and eastern Europe in the 1990s." J Epidemiol Community Health 60(10): 875-881.

Sullivan, D. F. (1971). "A single index of mortality and morbidity." HSMHA Health Rep 86(4): 347-354.

Van de Water, H. P., H. C. Boshuizen, et al. (1995). "Health expectancy: an indicator for change?" Journal of Epidemiology and Community Health 49(3): 330-331.
