

# The Groupwise Decomposition: Estimating Group-Specific Contributions to Life Table Functions

Eugenio Paglino<sup>1,2</sup>

1. Helsinki Institute for Demography and Population Health, University of Helsinki, Helsinki, Finland

2. Max Planck – University of Helsinki Center for Social Inequalities in Population Health (MaxHel), Helsinki (Finland) and Rostock (Germany)

## **Abstract**

**Background:** Several general decomposition algorithms have been developed to decompose the difference between the values of demographic measures in two populations. However, fewer techniques are available to researchers wishing to understand how different subpopulations are contributing to the difference in the value of a demographic measure between the total population and a reference subpopulation.

**Methods:** This paper proposes a general decomposition to answer these types of questions. As for other general decomposition techniques, the method assumes that the aggregate measure is computed from matrices of discrete demographic data but is agnostic to the specific function used to compute the measure. After proving some of the decomposition's properties, its relationships with other methods are discussed and extensions are considered.

**Results:** The proposed method is used to investigate the contribution of US counties by metropolitan/nonmetropolitan code to the life expectancy gap between large metropolitan areas and the US as a whole. Results from this new approach are compared to those obtained by adapting other general decomposition algorithms and their computational efficiency is evaluated.

**Contribution:** This paper provides a new decomposition method to investigate how different subpopulations contribute to the difference between a demographic measure for the total population and for a reference population. This setup is common in demography and the decomposition method developed in this paper adds one more tool to improve our understanding how different groups contribute to the demographic profile of a population.

## Introduction

The decomposition of demographic measures is a fundamental tool in a demographer's toolkit. Through the decomposition of demographic measures, researchers can gain crucial insights into the how different mechanisms produce demographic change. Kitagawa's classic decomposition method (Kitagawa 1955) aimed at decomposing the difference between the crude mortality rates of two populations into differences in age-specific mortality rates and differences in the age composition. Das Gupta generalized this approach to more than two dimensions (Das Gupta 1978), with the restriction that the aggregate demographic quantity had to be a linear function of the covariates of interest. In the 1980s the focus shifted to non-linear functions, and methods to decompose differences in life expectancy into age-specific components were independently developed by at least four scholars (Andreev 1982; Arriaga 1984; Pollard 1982; Pressat 1985). The next development involved designing general decomposition methods that could decompose a difference between two values of an arbitrary demographic index, with the requirement that it could be derived from a matrix of demographic quantities, into matrix element-specific components. Three such methods have been developed so far, two based on models for continuous change (Caswell 1989; Horiuchi, Wilmoth, and Pletcher 2008) and one based on a model for discrete, stepwise, change (Andreev, Shkolnikov, and Begun 2002).

None of these methods, nevertheless, are immediately suitable to tackling group-wise decomposition problems. As an example, consider a demographic measure for the total population  $f(T)$ , say US life expectancy, and the same demographic measure for a reference population  $f(R)$ , say life expectancy for the US-born population, and we wish to 1) understand how much individuals who are not in the reference population but are part of the total population, in our example the foreign-born population, contribute to  $f(T)$ , and 2) how their total contribution can be split into group-specific components, for example by race and Hispanic origin, by educational attainment, by sex, or any other dimension of interest. It turns out that the general decomposition techniques based on a continuous change model can be adapted to solve this decomposition challenge, but they quickly become computationally inefficient even when few groups are involved. This paper thus develops a

new decomposition technique which overcomes these limitations, discusses one practical application, and shows how the technique can be further extended.

## Relationship

Given a population ( $T$ ), a reference subpopulation ( $R$ ), and a partition of the remaining population into mutually exclusive subgroups ( $G_1, G_2, \dots, G_I$ ), indexed by  $i$ , such that  $D_x^R + \sum_{i=1}^I D_x^i = D_x^T$  and  $E_x^R + \sum_{i=1}^I E_x^i = E_x^{T^I}$ , how can we compute the contribution of subgroups  $G_1, G_2, \dots, G_I$  to the difference between the life table functions of the reference population ( $l_x^R, L_x^R, T_x^R, e_x^R$ ) and those of the total population ( $l_x^T, L_x^T, T_x^T, e_x^T$ )? More precisely, how can we find group-specific terms  $C^n$  that satisfy the following equation:

$$C^i \ (i = 1, 2, \dots, I) \quad s.t. \quad \sum_{i=1}^I C^i = f(T) - f(R) \quad (1)$$

Where  $f(\cdot)$  is a generic life table function which, as we will see, can be left unspecified. This paper proposes a decomposition method to answer this question and proves some of its properties. It then illustrates one practical application of the proposed decomposition.

## Strategy to Develop a Proof

The goal is to find a closed form expression for the  $C^i$  terms that satisfies Equation (1). I start by considering the simplest possible case, the one in which the population ( $T$ ) is partitioned into two subgroups, a reference population ( $R$ ), and a contributing group ( $G$ ). We can think of  $T$ ,  $R$ , and  $G$  as vectors of age-specific death counts and exposures stacked on each other. This is not essential for the proof but allows us to write expressions such as  $T = R + G$  with standard notation for operations between vectors. In the simplest case, it is trivial to show that the difference in the value that any life table function  $l(\cdot)$  takes for the total and the reference

---

<sup>1</sup> Here  $D_x^T$ ,  $D_x^R$ , and  $D_x^i$  are death counts for age  $x$  for the total population, the reference population, and the  $i^{th}$  residual subpopulation respectively. Similarly,  $E_x^T$ ,  $E_x^R$ , and  $E_x^i$  are the corresponding exposures in person-years.

population is equal to the contribution of the  $G$  subgroup or  $C^G = f(T) - f(R)$ . Starting from this basic equality, I will then move to the case of two contributing groups  $G_1$  and  $G_2$  and show that the expression for  $C^{G_1}$  (which from here onwards I will denote with  $C^1$ ) involves computing and averaging the differences between the value that the life table function takes for each combination of  $R$ ,  $G_1$ , and  $G_2$  that includes the target group  $G_1$  and the one that excludes it. Since the notation quickly becomes complicated, I will first introduce a more complex notation for addressing the case of three contributing groups  $I = 3$ , before finally moving to the general case of any  $I$ . In each case, the proof consists in showing that there exists a symmetric expression for each  $C^i$  term such that they collectively satisfy Equation (1). I demonstrate that the expression for each of the  $C^i$  involves computing a set of differences between the value taken by  $f(\cdot)$  for combinations of  $R$  and all  $G_j$  that include and exclude  $G_i$ , always asking what is the marginal contribution of adding the  $G_i$  group, and averaging them with appropriate weights.

## A Decomposition for the $I = 1$ and $I = 2$ Cases

Let us consider a population ( $T$ ) that is composed of a reference population ( $R$ ) and two mutually exclusive contributing groups ( $G_1, G_2$ ). Let us also denote the population combining ( $G_1, G_2$ ) as  $G$ , so  $G = G_1 + G_2$ . We can think of  $T$ ,  $R$ , and each of the  $G_i$ s as vectors stacking age-specific death counts and exposures. In standard demographic notation, we would write:

$$T = \begin{matrix} D_0 \\ D_1 \\ D_3 \\ \vdots \\ D_{110} \\ E_0 \\ E_1 \\ E_3 \\ \vdots \\ E_{110} \end{matrix}$$

Where  $D_x$  is the number of deaths to individuals aged  $x$  and  $E_x$  is the number of person-years lived by individuals aged  $x$ . We can then write  $G = T - R$ , which becomes a standard element-wise difference operation between vectors. At this point, I will be as general as possible and leave the life table function of interest unspecified. As in other general decomposition techniques, we will see that computationally all that matters is that we can write the life table measure as a function of a single input vector storing exposures and deaths counts (Andreev et al. 2002; Caswell 1989; Horiuchi et al. 2008; Riffe 2019). We will denote with  $l(A)$  the life table function for population  $A$ . To be precise, the generic  $f(\cdot)$  function takes as input a vector of age-specific death counts and exposures for population  $A$  (stacked to form a single vector) and outputs the life table measure  $f(A)$ . We can write the total contribution of the two groups as  $C^G = f(T) - f(R)$ . Because both the exposures and the death counts involved in the computation of life tables combine additively, we have that  $f(R) = f(T - G)$ . Equivalently, we could write  $f(T) = f(R + G)$ . Together, these two equations allow us to write:

$$C^G = f(T) - f(R) = f(R + G) - f(R) = f(T) - f(T - G) \quad (2)$$

Looking at the equation above it is apparent that the total contribution of population  $G$  can equivalently be understood as an addition process starting from the reference population,  $C^G = f(R + G) - f(R)$ , or as a subtraction process starting from the total population  $C^G = f(T) - f(T - G)$ . It now becomes useful to introduce another layer of notation. We define  $C^{G+} \equiv f(R + G) - f(R)$  and call it the *addition contribution* of  $G$ . We instead define  $C^{G-} \equiv f(T) - f(T - G)$  and call it the *subtraction contribution* of  $G$ . Clearly, when a single contributing subpopulation is considered  $C^{G+} = C^{G-}$ . However, we will see that even in the case of two contributing subpopulations  $(G_1, G_2)$ ,  $C^{1+} \neq C^{1-}$  and  $C^{2+} \neq C^{2-}$  so that the notation I just introduced becomes useful.

To see why these two inequalities arise, let us consider  $C^{1+}$  and  $C^{1-}$ . Suppose for the moment that both  $G_1$  and  $G_2$  contribute positively to the life table function we are considering, and more precisely that  $C^{1+}, C^{1-}, C^{2+}, C^{2-} > 0$ . When we compute  $C^{1-}$  we are then starting from a population  $T$  which already

encapsulates the positive effect of population  $G_2$ . On the contrary, when we compute  $C^{1+}$  we are starting from the reference population  $R$  that has not yet received the positive effect of population  $G_2$ . Furthermore, population  $R$  from which are starting in our calculations for  $C^{1+}$  is smaller than population  $R + G_2$  from which we start in our calculations for  $C^{1-}$ . Consequently, the effect of  $G_1$  will be larger when computed as  $C^{1+}$  than when computed as  $C^{1-}$  and  $C^{1+} > C^{1-}$ . A specular argument can be made for  $C^{2+}, C^{2-}$ . While the inequality  $C^{1+} > C^{1-}$  depends on the assumption that both populations have a positive effect, it illustrates the general principle that leads to the inequality  $C^{1+} \neq C^{1-}$ .

The inequality above means that we now have two different measures of the contribution of each group. One way of solving this inconsistency is to introduce a third type of contribution  $C^1 \equiv \frac{1}{2}(C^{1+} + C^{1-})$ , which we call the *average contribution*. It turns out that the average contributions have a nice property that both the addition and the subtraction contributions generally lack. In particular, we have:

$$\begin{aligned}
C^1 + C^2 &= \frac{1}{2}(C^{1+} + C^{1-}) + \frac{1}{2}(C^{2+} + C^{2-}) \\
&= \frac{1}{2}[(f(R + G_1) - f(R)) + (f(T) - f(T + G_1)) + (f(R + G_2) - f(R)) + (f(T) - f(T + G_2))] \\
&= \frac{1}{2}[2(f(T) - f(R)) + (f(R + G_1) - f(T - G_2)) + (f(R + G_2) - f(T - G_1))] \\
&= (f(T) - f(R)) + \frac{1}{2}[(f(R + G_1) - f(T - G_2)) + (f(R + G_2) - f(T - G_1))] \\
&= C^G + \frac{1}{2}[(f(R + G_1) - f(T - G_2)) + (f(R + G_2) - f(T - G_1))] \\
&= C^G + \frac{1}{2}[(f(R + G_1) - f(R + G_1)) + (f(R + G_2) - f(R - G_2))] \\
&= C^G
\end{aligned}$$

So, the two average contributions sum to the total contribution of  $G$ , which establishes  $C^1, C^2$  as a legitimate decomposition of  $C^G$  or of the difference  $f(T) - f(R)$ . Notice that no special properties of the life table function were used in the proof, which establishes that this approach is valid for functions other than life expectancy as long as they can be written as a function of a vector of death counts and exposures.

### Extending the Decomposition to the Case $I = 3$

The case of three groups  $(G_1, G_2, G_3)$  introduces a few additional complications which illuminate some properties of this decomposition and will allow us to develop a general expression for any number of groups. The main difference with the  $(I = 2)$  case is that our notation  $C^{i+}, C^{i-}$  for  $i = 1, 2, 3$  is no longer sufficient to describe all possible ways of computing the contribution of group  $G_i$  to the life table function for the total population. Indeed, we now have four ways of computing the contribution of  $G_1$ :

1.  $f(R + G_1) - f(R)$
2.  $f(R + G_2 + G_1) - f(R + G_2)$
3.  $f(R + G_3 + G_1) - f(R + G_3)$
4.  $f(T) - f(R + G_2 + G_3)$

Number 1 and 4 are equal to  $C^{1+}$  and  $C^{1-}$ , but number 2 and 3 are new expressions. In the next section, I will introduce a more general notation to capture these cases. However, for the moment, let us focus on how to define the average contribution in this case, with the idea of finding a set of terms for  $C^1, C^2$ , and  $C^3$  that sum to  $f(T) - f(R)$ . It turns out that to do so we need to construct a weighted average of the four different contribution types with weights inversely proportional to the number of groups  $I = 3$  and to the number of contributions involving the same number of subpopulations. Notice that contributions 1 and 4 involve 1 and 3 subpopulations respectively, while contributions 2 and 3 each involve 2 subpopulations. As such, in this weighting schema contributions 1 and 4 have weight  $\frac{1}{I} \frac{1}{1} = \frac{1}{3}$ , while contributions 2 and 3 have weight  $\frac{1}{I} \frac{1}{2} = \frac{1}{6}$ . The use of these weights is equivalent to first averaging within contributions involving the same



number of subpopulations and then averaging between contributions involving different numbers of subpopulations. While this might seem unintuitive, I will now show that these weights lead to a set of average contributions that sum to the total contribution as in the  $I = 2$  case.

$$\begin{aligned}
C^1 + C^2 + C^3 &= \frac{1}{3}(f(R + G_1) - f(R)) + \frac{1}{6}(f(R + G_2 + G_1) - f(R + G_2)) + \frac{1}{6}(f(R + G_3 + G_1) - f(R + G_3)) + \frac{1}{3}(f(T) - f(R + G_2 + G_3)) \\
&+ \frac{1}{3}(f(R + G_2) - f(R)) + \frac{1}{6}(f(R + G_1 + G_2) - f(R + G_1)) + \frac{1}{6}(f(R + G_3 + G_2) - f(R + G_3)) + \frac{1}{3}(f(T) - f(R + G_1 + G_3)) \\
&+ \frac{1}{3}(f(R + G_3) - f(R)) + \frac{1}{6}(f(R + G_1 + G_3) - f(R + G_1)) + \frac{1}{6}(f(R + G_2 + G_3) - f(R + G_2)) + \frac{1}{3}(f(T) - f(R + G_1 + G_2))
\end{aligned}$$

It is easy to see that one can take out of this complicated expression the target value  $f(T) - f(R)$ . We can then reorder the remaining terms and write:

$$\begin{aligned}
C^1 + C^2 + C^3 &= (f(T) - f(R)) \\
&+ \frac{1}{3}(f(R + G_1) + f(R + G_2) + f(R + G_3)) \\
&- \frac{1}{3}(f(R + G_1 + G_2) + f(R + G_1 + G_3) + f(R + G_2 + G_3)) \\
&- \frac{1}{6}(2 \cdot f(R + G_1) + 2 \cdot f(R + G_2) + 2 \cdot f(R + G_3)) \\
&+ \frac{1}{6}(2 \cdot f(R + G_1 + G_2) + 2 \cdot f(R + G_1 + G_3) + 2 \cdot f(R + G_2 + G_3))
\end{aligned}$$

Where the second and fourth lines cancel out and so do the third and fifth lines, leaving  $C^1 + C^2 + C^3 = f(T) - f(R)$  as we wanted to prove. This result establishes  $(C^1, C^2, C^3)$  as a legitimate decomposition of  $C^G$ .

### Generalizing the Decomposition to a Population with $I$ Subgroups

To study the general case of  $I$  contributing groups, it is useful to introduce a slightly more general notation. We define:

$$C^{i,m,k} = f(R + G_k^m + G_i) - f(R + G_k^m)$$

With  $i = \{1, 2, \dots, I\}$ ,  $m = \{0, 1, \dots, I - 1\}$ ,  $k = \{1, 2, \dots, \binom{I-1}{m}\}$  and where  $G_k^m$  is the  $k^{th}$  set of subpopulations  $G_j$  with  $j \neq i$  such that  $\#G_k^m = m$ . In words,  $C^{i,m,k}$  is the contribution of subgroup  $G_i$  to the life table function  $f(\cdot)$  for the total population calculated as the difference between the value of  $f(\cdot)$  for the population obtained by combining the reference population, the  $k^{th}$  subset of  $G_j$ 's not including  $G_i$ , and  $G_i$ ,  $f(R + G_k^m + G_i)$ , and the value of  $f(\cdot)$  for the population obtained by combining the reference population and the  $k^{th}$  possible subset of  $G_j$ 's not including  $G_i$ ,  $f(R + G_k^m)$ . This definition might seem confusing but it's just a generalization of the types of contributions we have seen in the cases  $I = 2, 3$ . Indeed, for  $I = 2$ , our familiar  $C^{1+}$  is simply  $C^{1,0,1}$  while  $C^{1-}$  is  $C^{1,1,1}$ . The key feature is that we are comparing the value a life table function takes for a population with and without the group whose contribution we want to estimate. Back to the  $I = 3$ , we can now write:

$$f(R + G_2 + G_1) - f(R + G_2) = C^{1,1,1}$$

$$f(R + G_3 + G_1) - f(R + G_3) = C^{1,1,2}$$

With these expressions, we can see why  $m$  should be bounded between 0 and  $I - 1$  and  $k$  between 1 and  $\binom{I-1}{m}$ . With  $m = 0$ , we recover  $C^{1+}$ , when  $m = 0$ , there is no need for an index  $k$ , which we can conventionally set to 1. For  $m = 1$ ,  $G_k^m$  contains only one element which we can choose out of the  $I - 1$  groups that are not  $G_i$ . The first of these groups will be denoted as  $G_1^1$ , the second as  $G_2^1$ , and so on. The order in which we select the groups does not matter because addition is commutative. Notice that, for a general value  $m$  there will be  $\binom{I-1}{m}$  of these groups, which explains the limits imposed on  $k$ . Finally, when we reach  $m = I - 1$ , we obtain  $C^{1-}$  because:

$$f(R + G_1^{I-1} + G_1) - f(R + G_1^{I-1}) = f(T) - f(T - G_1) = C^{1-}$$

Clearly, there is only one way of choosing  $I - 1$  elements from a set of  $I - 1$  elements. As for the  $I = 3$  case, we now need to find appropriate weights for each contribution  $C^{i,m,k}$  to construct  $C^i$  terms that sum to  $f(T) -$

$f(R)$ . The general principle is still the same, we need weights that are inversely proportional to the number of contribution types  $I$  and inversely proportional to the number of contributions for the specific type  $m$ . This consideration leads to the simple expression:

$$w^m = \frac{1}{I \binom{I-1}{m}} \quad (3)$$

Which is just a generalization of the weights we derived for ( $I = 3$ ). In the expression above,  $w^m$  denotes the weight for all terms  $C^{i,m,k}$ . These weights are equal for all  $m$ -type contributions and ensure that each set of  $m$ -contributions is collectively assigned the same weight  $\frac{1}{I}$  while each of its members is also assigned the same weight  $\frac{1}{I \binom{I-1}{m}}$ . With our new notation and having defined appropriate weights, we can define the average contribution of group  $G_i$  as:

$$C^i = \sum_{m=0}^{I-1} \frac{1}{I} \sum_{k=1}^{\binom{I-1}{m}} \frac{1}{\binom{I-1}{m}} C^{i,m,k} \quad (4)$$

To establish that average contributions defined in this way form a valid decomposition of  $f(T) - f(R)$  into group-specific contributions, we need to prove that:

$$\sum_{i=1}^I C^i = f(T) - f(R)$$

We already saw that for  $I = 2, 3$ , the proof involves showing that all terms except  $f(T)$  and  $f(R)$  have weights summing to 0. This is harder to do directly now that we have many terms. However, we can start by recognizing that the life table function computed for a given population composed by a set of subgroups plus the reference population can appear in just two ways:

1. As a term of the form  $f(R + G_k^m + G_i)$  in a contribution of the type  $C^{i,m,k}$
2. As a term of the form  $f(R + G_k^{m+1})$  in a contribution of the type  $C^{j,m+1,k}$  where  $j \neq i$ .

In the first case, the life table function will have weight  $w^m$  and in the second case it will have weight  $-w^{m+1}$ . To understand what the final weight will be for each term  $f(R + G_k^m + G_i)$ , we just need to know how many times it will appear with weight  $w^m$  and how many times with weight  $w^{m+1}$ . For terms of the first type, we have  $m + 1$  ways of choosing  $i$  while keeping the subgroups involved the same. On the other hand, terms of the second type can only appear in contributions involving the  $I - (m + 1)$  excluded groups. With this information, we are now able to compute the weight associated with the life table function computed for each population:

$$\begin{aligned}
w(l(R + G_k^m + G_i)) &= (m + 1)w^m - (I - m - 1)w^{m+1} \\
&= (m + 1) \frac{1}{I \binom{I-1}{m}} - (I - m - 1) \frac{1}{I \binom{I-1}{m+1}} \\
&= (m + 1) \frac{(I - 1)!}{m! (I - 1 - m)!} - (I - m - 1) \frac{(I - 1)!}{(m + 1)! (I - 1 - m - 1)!} \\
&= (m + 1) \frac{m! (I - 1 - m)!}{(I - 1)!} - (I - m - 1) \frac{(m + 1)! (I - 1 - m - 1)!}{(I - 1)!} \\
&= \frac{(m + 1)! (I - 1 - m)!}{(I - 1)!} - \frac{(m + 1)! (I - 1 - m)!}{(I - 1)!} \\
&= 0
\end{aligned}$$

Which proves that life table functions for all populations of the type  $R + G_k^m + G_i$  cancel out. The only two exceptions are  $f(R)$ , which appears  $I$  times with weight  $-\frac{1}{I}$ , and  $f(T)$ , which appears  $I$  times with weight  $\frac{1}{I}$ .

Thus, summing the average contributions for our  $I$  groups we obtain:

$$\sum_{i=1}^I C^i = f(T) - f(R)$$

as desired. This result shows that the average contributions provide a decomposition of the total contribution  $C^G$  even in the case of  $I$  subgroups.

## Origins of the Decomposition and Notes on Its Interpretation

The basic idea for the decomposition method developed in this paper was originally proposed by Hendi and Ho (2021) to investigate the contribution of the foreign born population to the national life expectancy at age 1 in the United States (Hendi and Ho 2021). In that context, Hendi and Ho considered a total population  $T$  including US residents, a reference population  $R$  including only US-born residents, and a single contributing group  $G$  including all foreign-born residents. They proposed to compute the contribution of the foreign-born population to the US life expectancy at age 1 as:

$$C^G = e_1^T - e_1^R$$

Where  $e_1^T$  is the life expectancy at age 1 for the total population and  $e_1^R$  is the life expectancy at age 1 for US-born residents (the reference population). The logic behind this formula is clear, the difference between the life expectancy of all US residents and US-born residents must be explained by the contribution of the foreign-born population. However, it is important to highlight some possible difficulties in the interpretation of such contribution before moving forward. Using Hendi and Ho (2021) example, if  $M_x^{USB}$  is the vector of period age-specific mortality rates for US-born residents,  $M_x^{FB}$  is the vector of period age-specific mortality rates for the foreign-born residents, and  $M_x^{US}$  is the vector of age-specific mortality rates for the total population, and assuming  $e_1^{FB} \geq e_1^{USB}$ , there is no guarantee that:

$$e_1^{USB} \leq e_1(M_x^{FB}w^{FB} + M_x^{USB}w^{USB}) = e_1^T \leq e_1^{FB}$$

Where  $w^{FB}$  and  $w^{USB}$  are vectors of age-specific weights proportional to the share of the foreign-born and US-born in the population, respectively, and such that  $w_x^{FB} + w_x^{USB} = 1 \forall x$ . The two inequalities fail to hold because a priori:

$$\max(e_1^T(M_x^{FB}w^{FB} + M_x^{USB}w^{USB})) \text{ w.r.t. } (w^{FB}, w^{USB}) = e_1(\min(M_x^{FB}, M_x^{USB}))$$

where  $\min(M_x^{FB}, M_x^{USB})$  is the set of elementwise (age-specific) minimums and:

$$\min(e_1^T(M_x^{FB}w^{FB} + M_x^{USB}w^{USB})) \text{ w.r.t. } (w^{FB}, w^{USB}) = e_1(\max(M_x^{FB}, M_x^{USB}))$$

where  $\max(M_x^{FB}, M_x^{USB})$  is the set of elementwise (age-specific) maximums. Equivalently, as long as the two sets of mortality rates  $M_x^{FB}, M_x^{USB}$  cross at some point (i.e.  $\exists i, j \text{ s.t. } M_i^{FB} > M_i^{USB} \wedge M_j^{FB} < M_j^{USB}$ ), there will be a set of weights such that  $e_1^T > \max(e_1^{FB}, e_1^{USB})$  and also a set of weights such that  $e_1^T < \min(e_1^{FB}, e_1^{USB})$ . The failure of  $e_1^T$  to be bounded by  $(e_1^{FB}, e_1^{USB})$  has two consequences that affect how we think of the contribution of a group  $G$  to the total life-expectancy given a certain reference population:

1. It can happen that  $C^G = e_1^T - e_1^R > e_1^G - e_1^R$ . In words, a group  $G$  could contribute more years to the total life expectancy than the difference between its life expectancy and that of the reference population.
2. Suppose we exchange the role of the contributing group  $G$  and that of the reference population  $R$ . We would then define:  $C^R = e_1^T - e_1^G$ . It is possible that both  $C^R > 0$  and  $C^G > 0$  even though  $R$  and  $G$ , by definition, form the total population  $T$ . Of course, it is also possible that  $C^R < 0$  and  $C^G < 0$ .

Given these two facts, it is important to understand that the contribution of group  $G$  computed with the Hendi and Ho (2020) formula, and similar contributions computed with the technique developed in this paper, depend on the mortality rates of the reference population and thus on the choice of a reference. More precisely, the contribution  $C^G$  does not capture the effect of group  $G$  “in isolation” but also its interaction with the mortality conditions of the reference population. Keeping this in mind, when the choice of a reference does not have strong theoretical foundations, researchers should test the robustness of the results to the choice of a different reference and, if major differences are found, discuss why it might be the case. While the choice of a reference may seem intuitive in the case of two populations, it will become less obvious when multiple groups are

considered, such as populations living in different Census Divisions, or in counties classified along the urban-rural continuum.

The extension of the Hendi and Ho (2021) expression to accommodate multiple subgroups follows the same idea as the stepwise decomposition (Andreev et al. 2002). The analogy is that we move from a vector with death counts and exposures for the reference population followed by a long vector of zeros to one where each set of zeros is replaced by the death counts and exposures for one of the  $G_i$ s, until all zeros are replaced, and we obtain the total population. The key difference between this group-wise replacement approach and the stepwise decomposition is that here the order in which the death counts and exposures for a group replace the zeros is highly consequential so that all permutations must be considered. Of course, approximations are possible and could speed-up the computation of the decomposition. However, simple strategies such as taking the average of the two *external* contributions  $C^{i+}$  and  $C^{i-}$  do not work well when the number of groups  $I$  is large (which is when we might want to use the approximation in the first place).

## Use Cases

The decomposition developed in this paper can be used to investigate the contribution of any set of mutually exclusive groups to the difference in the values of a life table function for the total and a reference population. As mentioned before, the basic decomposition for the  $I = 1$  case was introduced by Hendi and Ho (2021) to study how the foreign-born population contributes to the US life expectancy. However, the contributions subgroups of the foreign-born population, by origin or race/ethnicity or educational attainment for example, can be studied with the decomposition developed in this paper. Other interesting applications include the investigation of how different regions contribute to national life expectancy, how different countries contribute to the regional life expectancy, or how different educational groups contribute to the difference between the life expectancy of the group with the lowest educational attainment and that of the total population.

Like other general decompositions, while I have so far considered only examples related to mortality, applications to fertility, migration, or even projections are possible. Examples include computing the contribution of the foreign-born population to the national Total Fertility Rate (TFR) and decomposing it into education-specific terms. The same exercise could be carried out for the Net Reproduction Rate (NRR), combining fertility and mortality rates. In principle, even questions about future forecasted quantities can be addressed with this method. For example, we could use a Lee-Carter model (Lee and Carter 1992) to forecast mortality rates for the US as a whole and for large central metros in the US and ask how much counties in other metropolitan codes (fringe metros, medium metros, etc.) contribute to the difference between the two forecasts.

### **Example Application: The Contribution of Counties along the Urban-Rural Classification to the Life Expectancy Gap between Large Central Metros and All US Counties**

For this application, I use death counts and population data from the National Center for Health Statistics obtained through CDC WONDER. Deaths and population are classified by year (2017-2019), sex, five-year age groups (<1, 1-4, 5-9, ..., 85+), and one of six urban-rural codes. The six codes schema classifies counties into four metropolitan categories and two non-metropolitan categories. The metropolitan categories are large central metro, large fringe metro, medium metro, and small metro. The non-metropolitan categories are micropolitan and noncore. Details on the classification criteria are available in the NCHS online documentation (NCHS 2023).

As can be seen in Figure 1, at the national level, large fringe metro counties have the highest life expectancy for males, followed by large central metro counties, medium metro counties, small metro counties, micropolitan counties, and noncore counties. The ordering is essentially the same for females, the only difference being that large central metro counties have the highest life expectancy, with large fringe metro counties coming second. Figure 2 shows the group-specific contributions to national life-expectancy using the groupwise decomposition. As we would have expected, all groups but large fringe metro counties contribute



negatively to the national life expectancy. Micropolitan counties have the largest absolute contribution followed by noncore counties, medium metro counties, small metro counties, and large fringe metro counties. The size of each contribution is determined by the interaction of three factors: the population size of counties belonging to the specific group, the population's age distribution, and its age-specific mortality rates (which life expectancy conveniently summarizes).

As a further test of the validity of the groupwise decomposition, I adapted the three general decomposition methods (Andreev et al. 2002; Caswell 1989; Horiuchi et al. 2008) as implemented in the `DemoDecomp` package (Riffe 2019) to obtain a decomposition by group similar to the one developed in this paper. The results of the comparison between the results obtained with the groupwise decomposition and those obtained with each of the three methods are presented in Figure 3 and Figure 4 (in which the stepwise decomposition is removed). They show that the groupwise decomposition produces result almost identical to those of the line-integral method (Horiuchi et al. 2008) and close to the ones of the life table response experiment (Caswell 1989). The stepwise decomposition fails to produce reasonable results for the reason I outlined in the previous section.

A legitimate question is why one should use the groupwise decomposition if both the line-integral and the life table response experiment methods provide similar results. One answer is that the groupwise decomposition is computationally very efficient compared with the alternatives. For the example presented above, where a very simple life table function is used (`LTabr` from `DemoDecomp`) the average execution time measured with the `microbenchmark` package (Mersmann et al. 2023), was 0.15 seconds for the groupwise decomposition, 68.59 for the line-integral method, and 249.30 for the life table experiment, both of the latter as implemented in the `DemoDecomp` package. In this relatively simple application, the groupwise decomposition is thus about 460 times faster than the line-integral method and about 1,670 times faster than the life table experiment method. The computational advantage increases when more complex functions are used and more groups are considered.

## **A Simple Extension: Combining Decomposition Techniques to Compute Age- and Group-Specific Contributions**

Once we have seen how the decomposition works, we can quickly recognize two facts. First, while we only considered a function mapping vectors to scalars, nothing in the proof requires it. We can thus extend the decomposition to functions that take as input a vector of deaths counts and exposures and output a vector of contributions. Second, general decomposition techniques themselves are a useful class of vector-to-vector functions that satisfy this property. With these two realizations we see that we can easily combine other decomposition techniques with the groupwise decomposition to further decompose group-specific contributions by other dimensions.

Figure 5 shows the results of this extension, using age as the additional dimension. Figure 6 replicates the decomposition using the line-integral method and once again shows that the results are nearly identical. In this more complex application, the computational advantages of the groupwise decomposition are a bit smaller because the function being decomposed now outputs vectors rather than a scalar, while the computations for the line-integral decomposition remain the same. Again, using the `microbenchmark` package, I timed the mean execution time for the groupwise decomposition at 2.25 seconds compared with 94.77 seconds for the line-integral method as implemented in the `DemoDecomp` package (with the groupwise decomposition 42 times faster).

## **Conclusion**

This paper provides a new decomposition method to investigate how different subpopulations contribute to the difference between a demographic measure for the total population and for a reference population. This setup is not infrequent in demography and arises with research questions such as “how much do counties belonging to different groups in the metro-nonmetro continuum contribute to the large central metro mortality advantage compared with national rates?” or “how much do different educational groups contribute

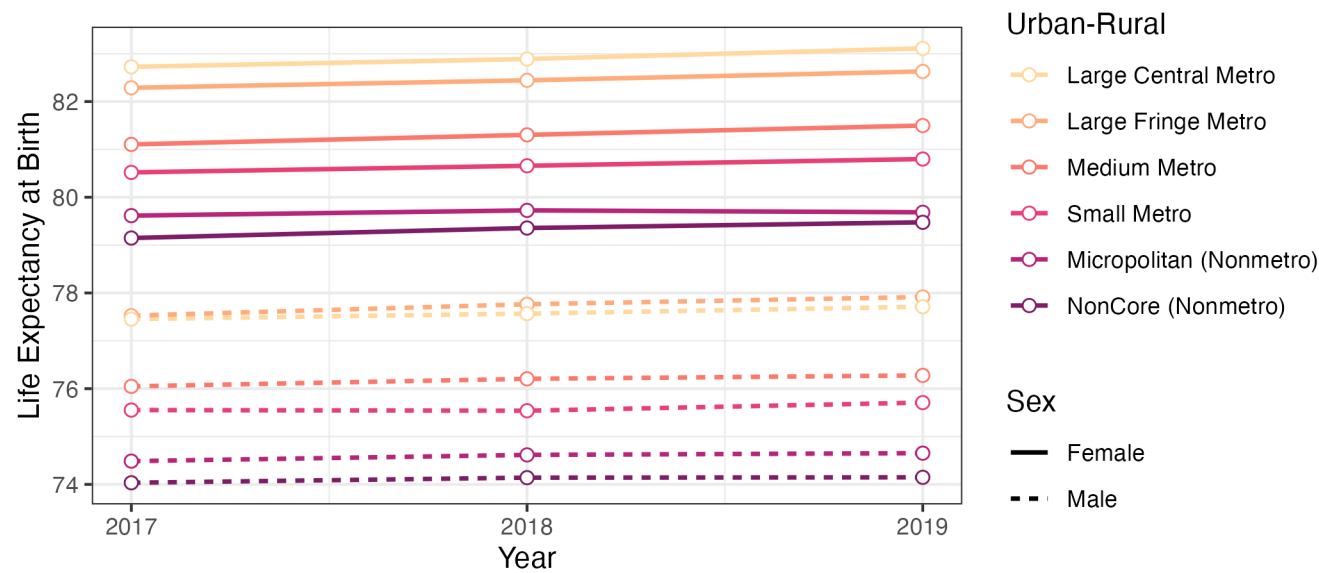
to the life expectancy gap between the population with a professional or graduate degree and the total population?” The decomposition method developed in this paper will enable researchers to answer these and other questions, expanding our understanding how different groups contribute to the demographic profile of a population.

## References

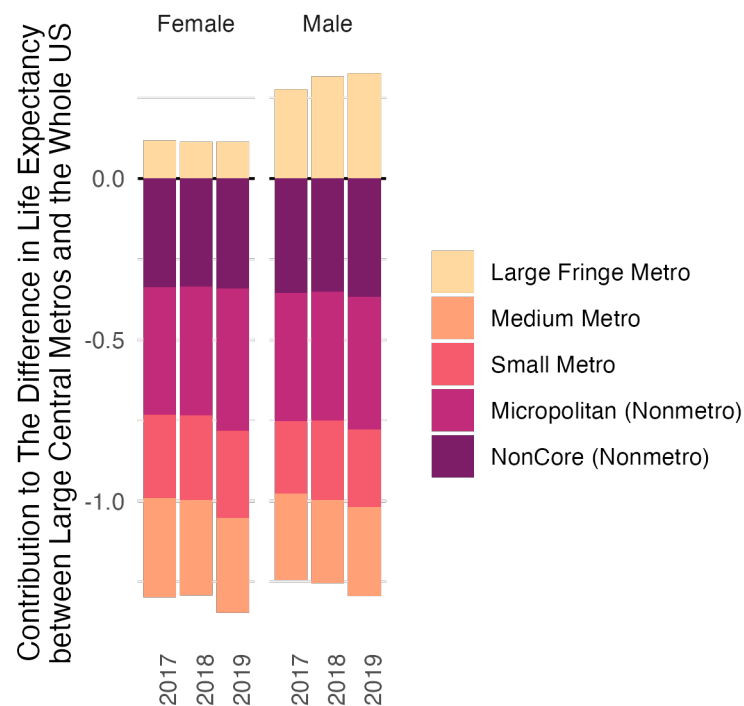
- Andreev, Evgueni M. 1982. "Metod Komponent v Analize Prodoljitelnosty Zjizni. [The Method of Components in the Analysis of Length of Life]." *Vestnik Statistiki* 9:42–47.
- Andreev, Evgueni M., Vladimir M. Shkolnikov, and Alexander Z. Begun. 2002. "Algorithm for Decomposition of Differences between 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, Eduardo E. 1984. "Measuring and Explaining the Change in Life Expectancies." *Demography* 21(1):83–96. doi: 10.2307/2061029.
- Caswell, Hal. 1989. "Analysis of Life Table Response Experiments I. Decomposition of Effects on Population Growth Rate." *Ecological Modelling* 46(3):221–37. doi: 10.1016/0304-3800(89)90019-7.
- Das Gupta, Prithwis. 1978. "A General Method of Decomposing a Difference between Two Rates into Several Components." *Demography* 15(1):99–112. doi: 10.2307/2060493.
- Hendi, Arun S., and Jessica Y. Ho. 2021. "Immigration and Improvements in American Life Expectancy." *SSM - Population Health* 15:100914. doi: 10.1016/j.ssmph.2021.100914.
- Horiuchi, Shiro, John R. Wilmoth, and Scott D. Pletcher. 2008. "A Decomposition Method Based on a Model of Continuous Change." *Demography* 45(4):785–801. doi: 10.1353/dem.0.0033.
- Kitagawa, Evelyn M. 1955. "Components of a Difference Between Two Rates." *Journal of the American Statistical Association* 50(272):1168–94. doi: 10.1080/01621459.1955.10501299.
- Lee, Ronald D., and Lawrence R. Carter. 1992. "Modeling and Forecasting U.S. Mortality." *Journal of the American Statistical Association* 87(419):659–71. doi: 10.1080/01621459.1992.10475265.
- Mersmann, Olaf, Claudia Beleites, Rainer Hurling, Ari Friedman, and Joshua M. Ulrich. 2023. "Microbenchmark: Accurate Timing Functions."
- NCHS. 2023. "NCHS Urban Rural Classification Scheme for Counties." Retrieved September 21, 2023 ([https://www.cdc.gov/nchs/data\\_access/urban\\_rural.htm](https://www.cdc.gov/nchs/data_access/urban_rural.htm)).
- Pollard, J. H. 1982. "The Expectation of Life and Its Relationship to Mortality." *Journal of the Institute of Actuaries* 109(2):225–40. doi: 10.1017/S0020268100036258.
- Pressat, Roland. 1985. "Contribution Des Écarts de Mortalité Par Âge à La Différence Des Vies Moyennes." *Population (French Edition)* 40(4/5):766–70. doi: 10.2307/1532986.
- Riffe, Tim. 2019. "DemoDecomp: General Demographic Decomposition Methods."

# Figures and Tables

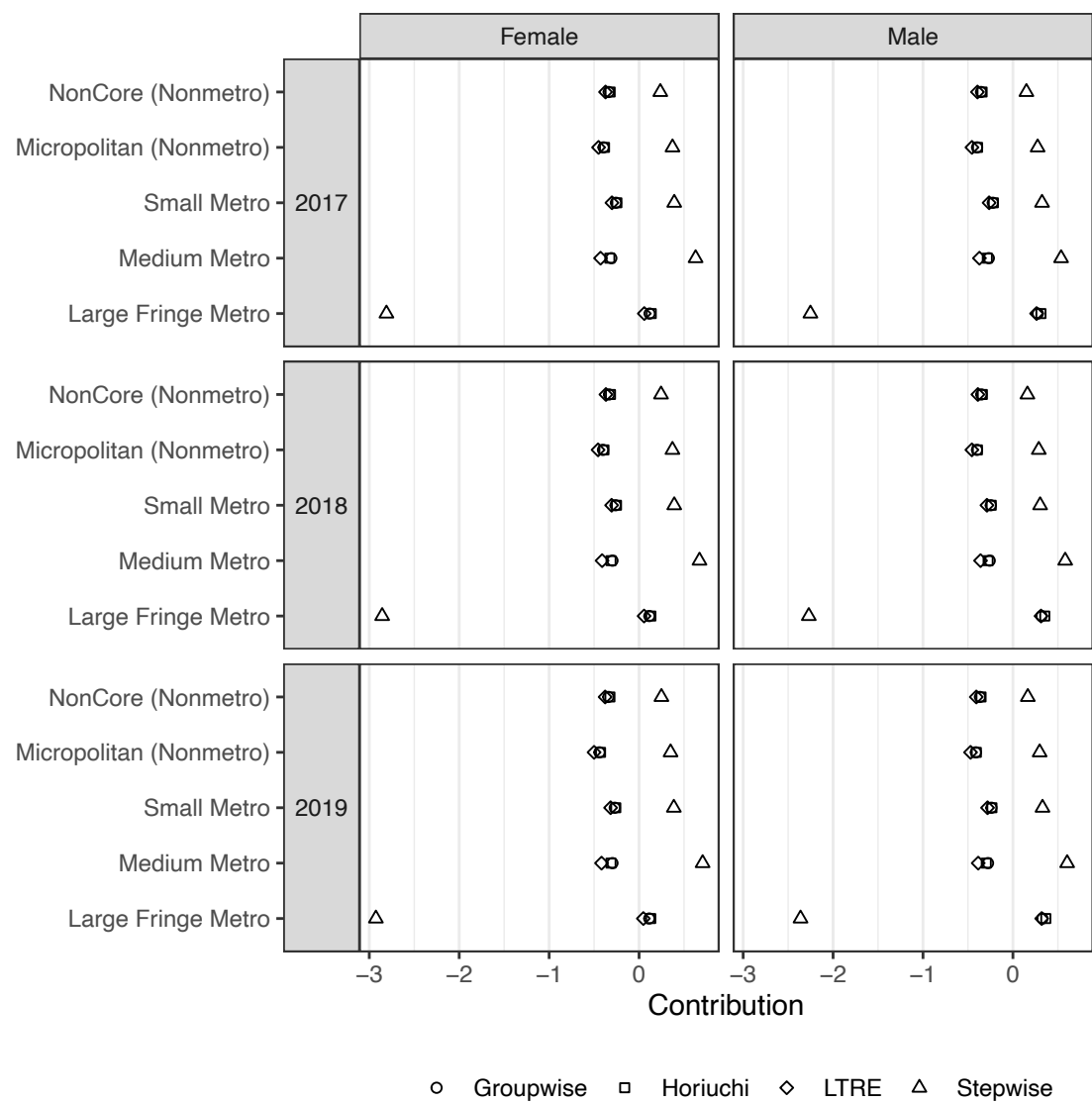
**Figure 1:** Life Expectancy at Birth by Year, Sex, and Urban-Rural Category



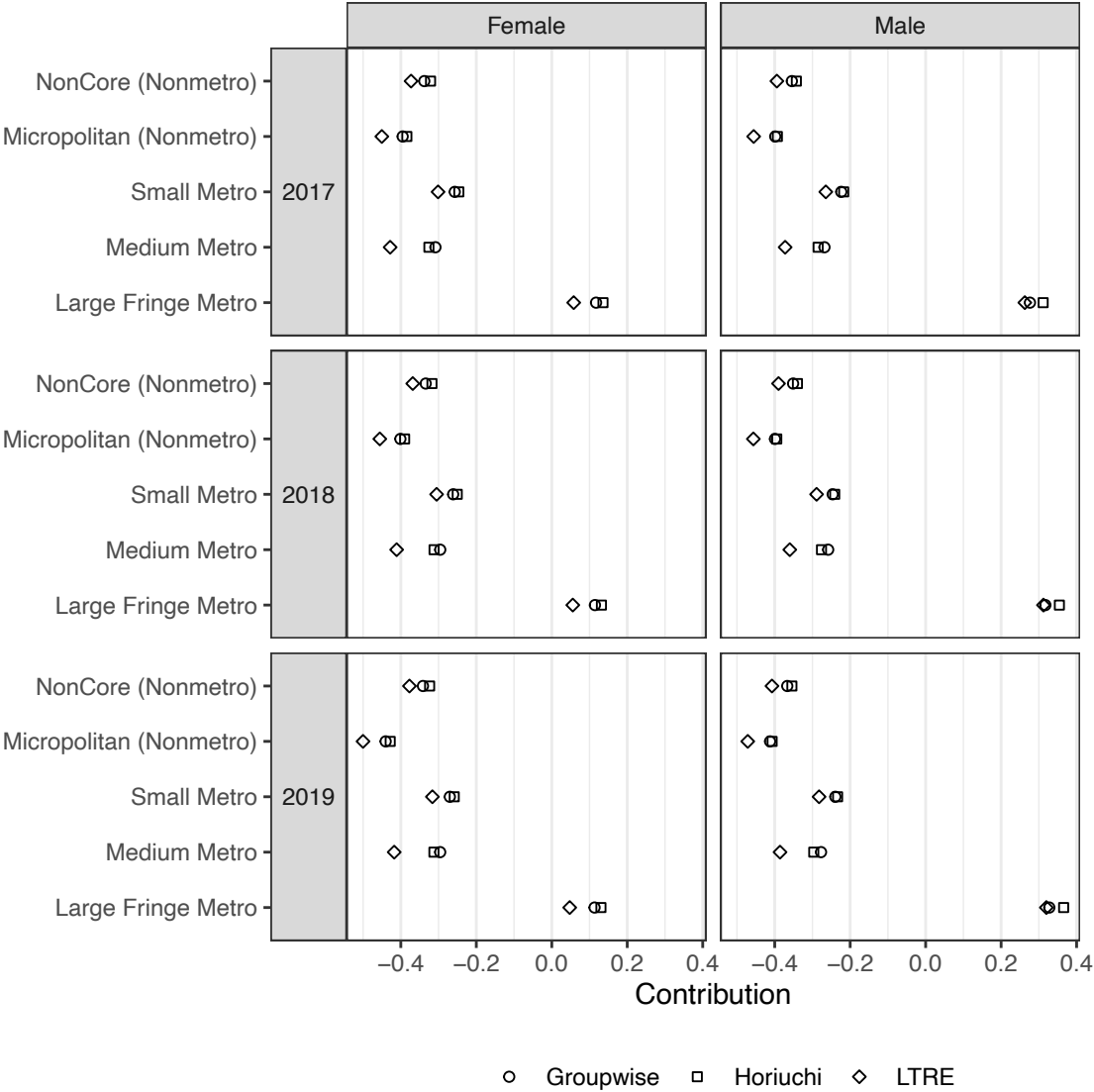
**Figure 2:** Decomposing the Contribution of Non-Large Central Metro Counties to National Life Expectancy by Urban-Rural Categories.



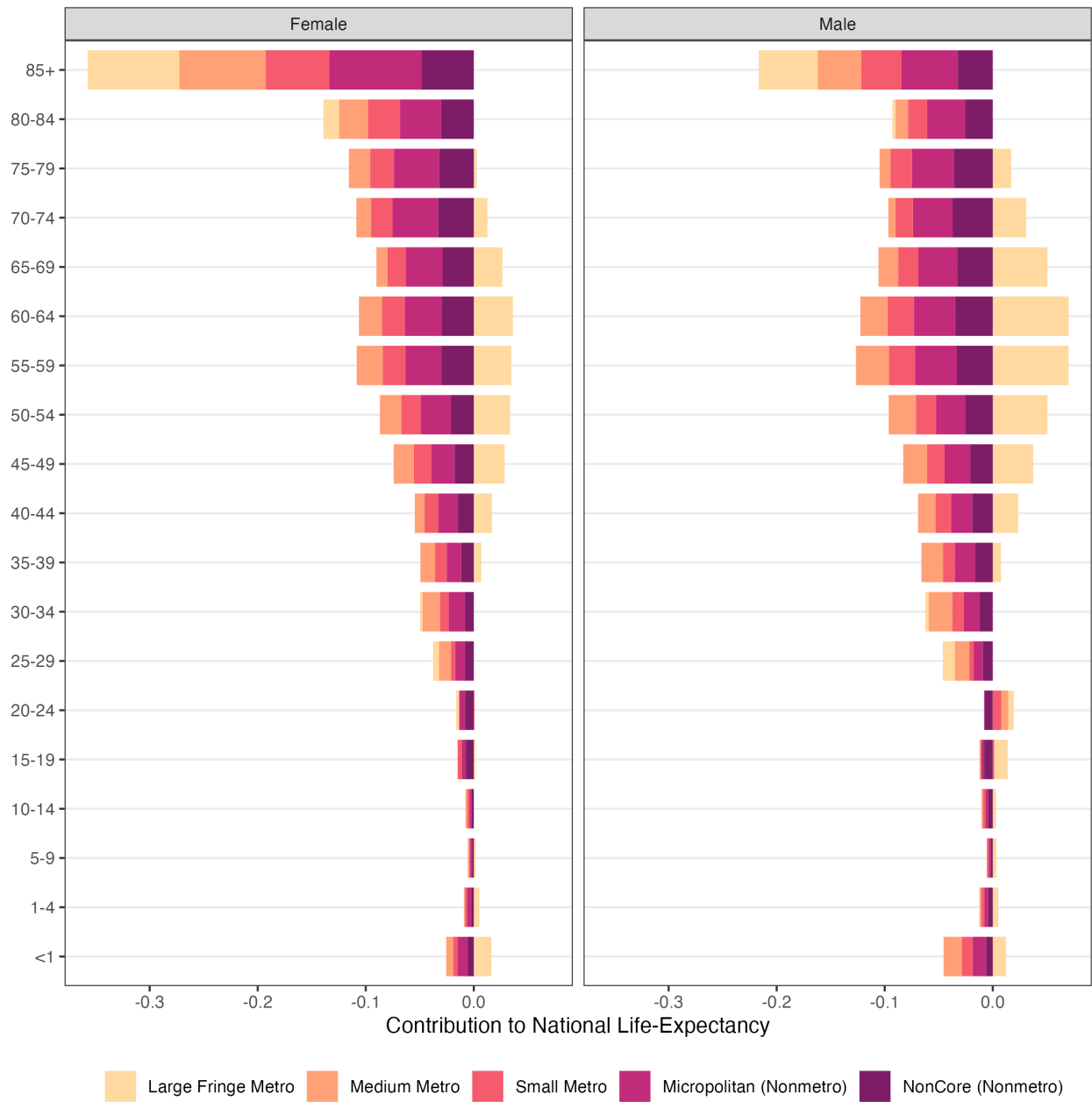
**Figure 3:** Comparing Decomposition Results from the groupwise decomposition with Those Obtained with Other General Decomposition Methods: Horiuchi, LTRE, and Stepwise.



**Figure 4:** Comparing Decomposition Results from the groupwise decomposition with Those Obtained with the Horiuchi and LTRE General Decomposition Methods.



**Figure 5:** Decomposition of Contributions to National Life Expectancy by Metro Type and Age (2019) Using the groupwise decomposition.





**Figure 6:** Decomposition of Contributions to National Life Expectancy by Metro Type and Age (2019) Using the Line-Integral Decomposition (Horiuchi et al. 2008).

