



# DEMOGRAPHIC RESEARCH

*A peer-reviewed, open-access journal of population sciences*

---

## ***DEMOGRAPHIC RESEARCH***

**VOLUME 39, ARTICLE 37, PAGES 991–1008  
PUBLISHED 26 OCTOBER 2018**

<http://www.demographic-research.org/Volumes/Vol39/37/>

DOI: 10.4054/DemRes.2018.39.37

*Formal Relationship 28*

**Life lived and left: Estimating age-specific  
survival in stable populations with unknown  
ages**

**James W. Vaupel**

**Francisco Villavicencio**

© 2018 James W. Vaupel & Francisco Villavicencio.

*This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit.*

*See <https://creativecommons.org/licenses/by/3.0/de/legalcode>*

## Contents

|     |   |      |
|-----|---|------|
| 1   | Relationship  | 992  |
| 2   | Proof   | 992  |
| 3   | Related results                                       | 993  |
| 3.1 | The stable population model                           | 994  |
| 3.2 | Life lived and left in stationary populations         | 995  |
| 3.3 | The population of France from a life left perspective | 997  |
| 4   | Application   | 997  |
| 4.1 | Life lived and left in a discrete-time framework      | 998  |
| 4.2 | Example of application with simulated data            | 1000 |
| 5   | The 1805 cohort life table of Swedish females         | 1002 |
| 5.1 | Estimation of the growth rate                         | 1003 |
| 5.2 | Reconstruction of the survival curve                  | 1003 |
| 6   | Discussion  | 1005 |
|     | References  | 1006 |

## **Life lived and left: Estimating age-specific survival in stable populations with unknown ages**

**James W. Vaupel<sup>1</sup>**

**Francisco Villavicencio<sup>2</sup>**

### **Abstract**

#### **BACKGROUND**

Demographers sometimes observe remaining lifespans in populations of individuals of unknown age. Such populations may have an age structure that is approximately stable. To estimate life tables for these populations, it is useful to know the relationship between the number of individuals at a given age  $a$ , and the number of individuals that are expected to die  $a$  time units after observation. This result has already been described for stationary populations, but here we extend it to stable populations.

#### **RESULTS**

In a stable population, the population at a given age  $a$  is a simple function of the number of deaths at remaining lifespan  $a$ , the number of deaths at remaining lifespans  $a$  and higher, and the population growth rate. This property, which can be useful when ages are unknown, but individuals are followed until death, permits estimation of the underlying unknown survival schedule of the population and calculation of the usual life table functions and statistics, including the stable age structure.

#### **CONTRIBUTION**

The main contribution of this article is to provide a formal proof of the relationship between life lived and life left in stable populations. We also discuss the challenges of applying theoretical relationships to empirical data, especially due to the fact that in real-world applications time is not continuous and some adjustments are necessary to move from a continuous to a discrete-time framework. Two applications, one with simulated data and a second with Swedish data from the 19<sup>th</sup> century, illustrate these ideas.

---

<sup>1</sup> Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Odense, Denmark.  
Email: [jvaupel@sdu.dk](mailto:jvaupel@sdu.dk).

<sup>2</sup> Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Odense, Denmark.  
Email: [fvillavicencio@health.sdu.dk](mailto:fvillavicencio@health.sdu.dk).

## 1. Relationship

Consider a stable population with constant growth rate  $r$  that is first observed at time  $t$ . Let  $N(a, t)$  be the unknown number of individuals in this population at any age  $a$ . Let  $\tilde{D}(a, t)$  be the number of individuals in the population at time  $t$  who die  $a$  time units after first observation. These are the individuals at time  $t$  who have exactly  $a$  time units of life left. The term ‘life left’ pertains to the remaining lifetime until death, as opposed to ‘life lived’ or age of an individual or cohort.<sup>3</sup> Let  $\tilde{N}(a, t) = \int_a^\infty \tilde{D}(x, t) dx$ : This function gives the number of individuals from the initial population who are still alive at time  $t + a$ . Then, we have that

$$(1^*) \quad N(a, t + a) = \tilde{D}(a, t) + r \tilde{N}(a, t),$$

where  $N(a, t + a)$  is the number of individuals at age  $a$  at time  $t + a$ .<sup>4</sup> Furthermore, given a population in which ages are unknown but individuals are followed until death, this result can be used to derive the underlying unknown cohort survival schedule,

$$(2^*) \quad \ell(a) = \frac{N(a, t + a)}{N(0, t)} = \frac{\tilde{D}(a, t) + r \tilde{N}(a, t)}{\tilde{D}(0, t) + r \tilde{N}(0, t)}.$$

## 2. Proof

Stable population theory implies that

$$\tilde{D}(x, t) = \int_0^\infty B(t) e^{-ra} d(a + x) da = \int_x^\infty B(t) e^{-r(a-x)} d(a) da,$$

where  $B(t)$  are the births at time  $t$  and  $d(a + x)$  is the unconditional risk of death at age  $a + x$ . Applying the Leibniz rule for differentiation under the integral sign, we get

$$\begin{aligned} \frac{d}{dx} \tilde{D}(x, t) &= \frac{d}{dx} \int_x^\infty B(t) e^{-r(a-x)} d(a) da = \lim_{M \rightarrow \infty} \frac{d}{dx} \int_x^M B(t) e^{-r(a-x)} d(a) da \\ &= -B(t) d(x) + \lim_{M \rightarrow \infty} r \int_x^M B(t) e^{-r(a-x)} d(a) da \\ &= -B(t) d(x) + r \tilde{D}(x, t), \end{aligned}$$

<sup>3</sup> Other terms such as ‘time-to-death,’ ‘remaining lifespan,’ ‘follow-up duration,’ or ‘residual life’ are also frequently used to refer to life left.

<sup>4</sup> All the main results not previously published are marked with an asterisk next to the equation number.

and, rearranging terms,

$$(3) \quad B(t) d(x) = -\frac{d}{dx} \tilde{D}(x, t) + r \tilde{D}(x, t).$$

Because  $\ell(a) = \int_a^\infty d(x) dx$  for all  $a \geq 0$ , integrating both sides of (3) yields

$$\begin{aligned} B(t) \ell(a) &= \int_a^\infty -\frac{d}{dx} \tilde{D}(x, t) + r \tilde{D}(x, t) dx \\ &= \lim_{M \rightarrow \infty} \int_a^M -\frac{d}{dx} \tilde{D}(x, t) dx + r \int_a^\infty \tilde{D}(x, t) dx \\ &= -\lim_{M \rightarrow \infty} \tilde{D}(M, t) + \tilde{D}(a, t) + r \tilde{N}(a, t) \\ &= \tilde{D}(a, t) + r \tilde{N}(a, t). \end{aligned}$$

Since  $\ell(a)$  is the probability of surviving from birth to age  $a$  and  $B(t)$  are the births at time  $t$ , it follows that  $B(t) \ell(a) = N(a, t + a)$ , the population at age  $a$  at time  $t + a$ , which proves (1\*). Equation (2\*) follows directly from (1\*) and the definition of the cohort survival function  $\ell(a)$ .

### 3. Related results

The main result in (1\*) that describes the relationship between life lived and life left in stable populations has already been proved for the particular case of stationary populations with growth rate  $r = 0$  (Brouard 1989; Vaupel 2009; Villavicencio and Riffe 2016). To our knowledge, Vaupel was the first to generalize it to stable populations in an unpublished manuscript from 2013. Here we polish Vaupel’s proof and include some additional material.

In the stable population model, the crude birth rate  $b$  and the crude death rate  $d$  are assumed to be constant over time, which permits estimation of the constant growth rate  $r$  of the population. Let  $N(t)$ ,  $B(t)$  and  $D(t)$  denote, respectively, the population size, the number of births, and the number of deaths at time  $t$  in a stable population. Then

$$(4) \quad r = b - d = \frac{B(t)}{N(t)} - \frac{D(t)}{N(t)},$$

and

$$(5) \quad B(t) = D(t) + r N(t).$$

Equation (5) can be re-expressed as follows: In stable populations, at any time  $t$  the number of individuals with 0 life lived (births) equals the number of individuals with

0 life left (deaths), plus an extra term that only depends on the constant growth rate and the population size. Equation (1\*) extends this relationship to estimate the number of individuals at older ages when the ages are unknown but there is information about the remaining time until death. In population biology, for instance, one may think of a wild population that is captured at a certain time point, and then the deaths and the number of individuals that are still alive at each subsequent time point are recorded until extinction. This information can be used to estimate the age-specific survival, provided that the assumption of stability is not overly distorting and that it is possible to estimate the population growth rate  $r$ .

Equations (4) and (5) pertain to a population with an initial age of 0, but they can be generalized to apply to a population with any initial age  $a$ . Then, at time  $t$ ,  $B(t)$  would be the number of individuals who are age  $a$ ,  $D(t)$  the number of individuals who die at age  $a$  and older, and  $N(t)$  the number of individuals alive at age  $a$  and older. The growth rate  $r$  also pertains to the population above age  $a$ , but in the stable model this  $r$  is the same for the entire population. Hence, Equation (1\*) can be viewed as a generalization of (5). Note that  $N(0, t + 0) = B(t)$  because the population at age 0 at time  $t$  equals the births at time  $t$ . Further, the individuals who at time  $t$  have 0 remaining lifetime are the deaths, and therefore  $\tilde{D}(0, t) = D(t)$ . Finally, the individuals from the initial population at time  $t$  who are still alive at  $t + 0$  are the same, so  $\tilde{N}(0, t) = N(t)$ . As a result, (5) leads to (1\*) when  $a = 0$ :

$$B(t) = D(t) + r N(t) \implies N(0, t) = \tilde{D}(0, t) + r \tilde{N}(0, t).$$

### 3.1 The stable population model

In 1760 Leonhard Euler (1707–1783) was the first to formally describe the concept of a stable population, defined as a population closed to migration experiencing fixed age-specific death rates, and births that vary in geometric progression for a prolonged period of time (Euler 1970). Euler’s work remained rather unknown to the scientific community, and his ideas were published during subsequent decades and centuries by scholars who independently rediscovered them. The development of a fully articulated theory of stable populations was due to Alfred J. Lotka (1880–1949) in a series of more than 30 articles and a book, *Théorie analytique des associations biologiques*. This book was first published in two separate volumes in 1934 and 1939 and, surprisingly, not translated into English until 1998 (Lotka 1934, 1939, 1998).

In stable populations, the population size at time  $t$  is given by

$$N(t) = \int_0^\infty N(a, t) da = \int_0^\infty B(t) e^{-ra} \ell(a) da$$

where  $B(t) e^{-ra}$  are the births at  $t - a$ . This definition implies that the population at any

given age at time  $t$  is  $N(a, t) = e^{-ra} N(a, t + a)$ . Therefore, our basic result (1\*) can also be used to estimate the unknown age structure at initial time  $t$ ,

$$(6^*) \quad c(a, t) = \frac{N(a, t)}{N(t)} = \frac{e^{-ra} N(a, t + a)}{N(t)} = e^{-ra} \frac{\tilde{D}(a, t) + r \tilde{N}(a, t)}{N(t)}.$$

It may also be interesting to observe that (6\*) is equivalent to a more common expression for the proportion of the population at age  $a$  in terms of the birth rate  $b$ ,

$$(7) \quad c(a, t) = b e^{-ra} \ell(a),$$

as stated by Lotka more than a century ago (Lotka 1907). Equation (7) implies that the population structure does not depend on time because the birth rate  $b$  is constant. Here, we use the notation  $c(a, t)$  to highlight that we refer to the age structure at the time of capture when the follow-up of the population starts. But in stable populations, the period age structure – as well as the cohort survival schedule – is constant over time, so one could simply write  $c(a)$ .<sup>5</sup>

Although stability is an abstract mathematical concept resulting from the continuous operation of strong demographic assumptions over the long run, these restrictions have turned out to be serviceable in many applications. Stable population theory has been widely used in the study of human and nonhuman populations, especially when accurate data is incomplete or problematic, as discussed in the manual *The Concept of a Stable Population: Application to the Study of Populations of Countries with Incomplete Demographic Statistics* (United Nations 1968). Even though perfect stability is rarely observed, this book argues that many populations around the world are ‘semi-stable’ or ‘quasi-stable,’ and possess some of the properties of the stable population model. Semi-stable refers to populations that have an unchanging age structure, whereas quasi-stable describes populations in which fertility remains unchanged and mortality improves gradually.

### 3.2 Life lived and left in stationary populations

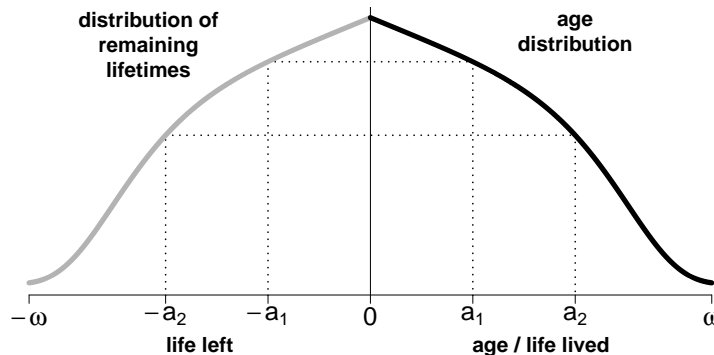
In stationary populations, in addition to fixed age-specific death rates and closure to migration, the birth flow is constant over time because of the growth rate  $r = 0$ , which implies that, in the long run, the number of births equals the number of deaths and the population size remains constant. This approach is more restrictive than stability but can lead to new insights about population dynamics. Some of these insights pertain to the symmetries between life lived and life left. Kim and Aron (1989) – and later Goldstein

<sup>5</sup> For a general overview of the main properties of stable populations, see for instance Preston, Heuveline, and Guillot (2001, Chapter 7).

(2009) – proved that the average age of a stationary population equals average remaining life expectancy, whereas Riffe (2015) addressed the equivalence between the force of mortality by age and the force of increment by life left. Moreover, stationary populations present a symmetry between their age composition and the distribution of remaining lifetimes. This relationship was first described by Brouard (1986, 1989), and was later independently formulated by James Carey and colleagues in the study of the survival patterns of captive and follow-up cohorts of medflies (Müller et al. 2004, 2007). Vaupel (2009) proved that in stationary populations of infinite size and in continuous time, the probability that a randomly selected individual is age  $a$  equals the probability that an individual has exactly  $a$  time left until death. Villavicencio and Riffe (2016) suggested an alternative proof for empirical and finite stationary populations in a discrete-time framework.

An alternative perspective is to consider the births and deaths in stationary populations, which need to be continuously in balance to keep the population size constant. Hence, the number of individuals at age 0 (births) equals the number of individuals with 0 life left (deaths). Note that the first group forms a birth cohort, whereas in the second there are individuals from different ages who share the same time of death, forming a ‘death cohort.’ The key point is that this result is true for all ages when the population is stationary, meaning that the number of individuals at any age  $a$  equals the number of individuals with  $a$  life left within the population. Figure 1 illustrates these ideas: Life left is plotted in the negative horizontal axis to convey the notion of a ‘countdown.’

**Figure 1: Symmetries between the age composition and the distribution of remaining lifetimes in stationary populations**



Note: Parameter  $\omega$  refers to the maximum observed lifespan.

Equation (1\*) extends this result to stable populations, so it can be used with less restrictive assumptions. In this case there is not an exact symmetry between life lived



and left because of the growth rate  $r$ , but there is still a relationship between the age composition (captured by  $N(a, t)$ ) and the distribution of remaining lifetimes (captured by  $\tilde{D}(a, t)$ ). Note that if  $r = 0$ , then  $N(a, t + a) = N(a, t)$  and (1\*) becomes

$$N(a, t) = \tilde{D}(a, t).$$

This equivalence formalizes the concept of symmetry between life lived and left in stationary populations that has just been described: The number of individuals aged  $a$  at time  $t$  equals the number of individuals who will die at time  $t + a$ .

### 3.3 The population of France from a life left perspective

Brouard (1986) presents an original overview of the population dynamics of France in the 20th century from a life left perspective. He compares the age structure of the French population in several years with what he calls the ‘pyramide des années à vivre,’ that is, the projected structure of remaining lifetimes of the cohorts alive in those years. He also introduces the concept of ‘génération de décès’ (death cohort) mentioned above. Especially relevant are the two pyramids from 1901, which show that the male population above age 70 is surprisingly close to the population that at that time had more than 70 years of life left and lived at least until 1970. Lower remaining lifetimes were caused by traumatic events such as the two world wars that increased mortality, but in general there is a fairly close symmetry between both pyramids. In contrast, in the case of women the improvements in mortality were more marked, and thus the pyramid of remaining lifetimes has more individuals with higher values, given that more women lived longer than 70 years after 1901 than those above age 70 in that year. Brouard concludes that the analysis of the remaining lifetimes offers a new interesting perspective for the study of population dynamics, notably for older ages. He suggests the estimation of life left according to some health markers, which could provide better insights into the characteristics of the old population and help in the assessment of proper policies.

## 4. Application

Demographers like to think of time as a continuous process, which permits the use of differential calculus and facilitates the proof of demographic relationships among demographic functions. This has been the case in the statement and proof of (1\*) and (2\*), as well in the works mentioned above that are based on the stable population theory developed by Lotka (1939).

In a theoretical population model, it is reasonable to assume time to be continuous, and compute vital rates and the instantaneous – although theoretical and potentially

noninteger – number of births and deaths at any time. On the contrary, in an empirical population in which individuals are indivisible, the instantaneous number of births and deaths would be zero, and the number of births and deaths would have to be computed in time intervals between observations. Villavicencio and Riffe (2016) adopted this approach to provide an alternative proof of the symmetries between life lived and left in empirical stationary populations. This approach represents a more realistic setup, but has the disadvantage of making the demonstration more complex than in the continuous case.

#### 4.1 Life lived and left in a discrete-time framework

In an empirical setup, some adjustments are necessary to adapt Equations (1\*), (2\*), and (6\*) to a discrete-time framework. In general, we can only get the population counts at exact times  $t, t + 1, t + 2$ , etc., and we may not be able to know what births and deaths occurred between observations. Moreover, a slightly different notation is usually adopted to account for the fact that ages are not exact: In the following, we use subindices for the age variable to refer to the ‘age interval  $[a, a + 1)$ ’ instead of the ‘exact age  $a$ .’

Suppose a population is first captured at time  $t$  and the total number of individuals  $N(t)$  is recorded. This same population is observed again at time  $t + 1$  and the number of individuals still alive are recorded, obtaining  $\tilde{N}_1(t)$ , the individuals from the initial population who are alive at time  $t + 1$ , all of them at age 1 and above. The assumption of stability implies

$$(8) \quad N_{1+}(t) = e^{-r} \tilde{N}_1(t),$$

where  $N_{1+}(t)$  is the unknown number of individuals that at exact time  $t$  were age 1 and above. Equation (8) can be used to estimate the population at age  $[0, 1)$  at exact time  $t$ :

$$\begin{aligned} N_0(t) &= N(t) - N_{1+}(t) \\ &= \tilde{N}_0(t) - e^{-r} \tilde{N}_1(t), \end{aligned}$$

given that  $N(t) = \tilde{N}_0(t)$ , the population at age 0 and above at exact time  $t$ . Analogously,

$$\begin{aligned} N_1(t + 1) &= \tilde{N}_1(t) - N_{2+}(t + 1) \\ &= \tilde{N}_1(t) - e^{-r} \tilde{N}_2(t), \end{aligned}$$

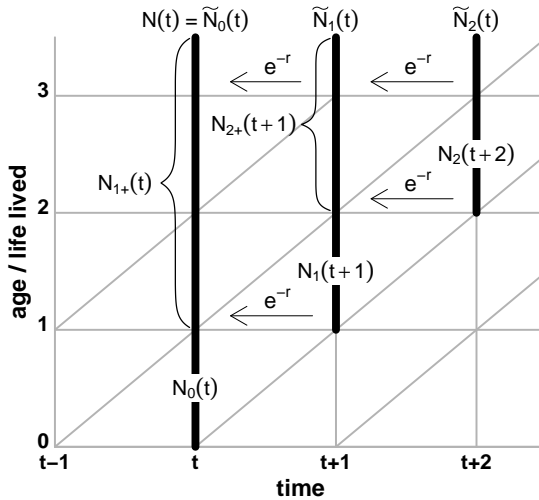
where  $N_1(t + 1)$  is the number of individuals at age  $[1, 2)$  at exact time  $t + 1$ ,  $N_{2+}(t + 1)$  is the number of individuals at age 2 and above at exact time  $t + 1$ , and  $\tilde{N}_2(t)$  is the number of individuals from the initial population who are still alive at exact time  $t + 2$ . This result can be extended to any age  $a$ , such that the population at age  $[a, a + 1)$  at exact

time  $t + a$  is

$$(9^*) \quad N_a(t + a) = \tilde{N}_a(t) - e^{-r} \tilde{N}_{a+1}(t).$$

Equation (9\*) is an extension of our basic result (1\*) to the case of observations made not continuously but at points in time  $t$ ,  $t + 1$ ,  $t + 2$ , etc., and is valid for any time unit – minutes, hours, days, years, or decades. It is necessary, though, that the same unit is used to measure time and age, and that the population is observed at regular time points. The Lexis diagram in Figure 2 provides some intuition about these concepts and relationships.

**Figure 2:** Lexis diagram illustrating the relationship between life lived and life left in a discrete-time framework



Note: Life lived is captured by  $N(t)$ ,  $N_0(t)$ ,  $N_1(t + 1)$ , etc., and life left by  $\tilde{N}_0(t)$ ,  $\tilde{N}_1(t)$ ,  $\tilde{N}_2(t)$ , etc. following Equation (9\*).

An equivalent expression to the survival schedule in (2\*) for the discrete case follows immediately from (9\*),

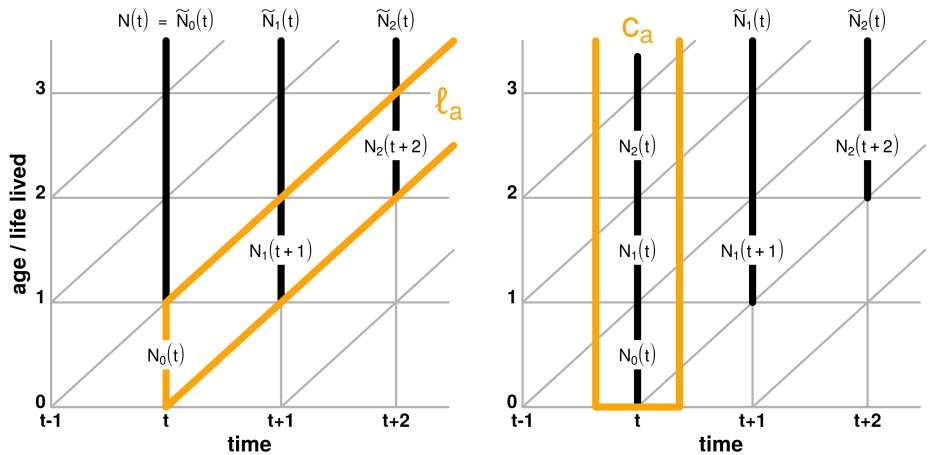
$$(10^*) \quad \ell_a = \frac{N_a(t + a)}{N_0(t)} = \frac{\tilde{N}_a(t) - e^{-r} \tilde{N}_{a+1}(t)}{\tilde{N}_0(t) - e^{-r} \tilde{N}_1(t)},$$

and the stable age structure defined in (6\*) becomes

$$(11^*) \quad c_a = \frac{N_a(t)}{N(t)} = \frac{e^{-ra} N_a(t+a)}{N(t)} \\ = e^{-ra} \frac{\tilde{N}_a(t) - e^{-r} \tilde{N}_{a+1}(t)}{N(t)}.$$

Figure 3 illustrates these two last results in a Lexis diagram: By following a stable population with unknown ages from time  $t$  onward, and recording at each subsequent time step the number of remaining individuals alive ( $\tilde{N}_0(t)$ ,  $\tilde{N}_1(t)$ ,  $\tilde{N}_2(t)$ , etc.), we are able to recover the cohort survival schedule  $\ell_a$  (left panel) and the stable age structure  $c_a$  (right panel).

**Figure 3:** Lexis diagrams illustrating the cohort survival schedule  $\ell_a$  (left panel) and the stable age structure  $c_a$  (right panel) that can be estimated by following a stable population until extinction



Note: Stable population with unknown ages that is followed from time  $t$  onward, recording at each subsequent time step the number of remaining individuals alive, and applying (10\*) and (11\*).

## 4.2 Example of application with simulated data

To check the validity of our results, we proceed with a reproducible example based on simulating a stable population. Following the same idea that James Carey and colleagues developed in the study of survival patterns of captive medflies (Müller et al. 2004, 2007),

we observe a population at regular time points until its extinction, and we record the  $\tilde{N}_a(t)$  values at each time step. Applying (9\*), (10\*), and (11\*), we use these values to estimate the survival schedule used in the data simulation and the corresponding stable age structure. The R code (R Core Team 2018) to carry out this experiment is available in the supplementary materials; only a few details are provided below.

We adapt the same simplified example of a stable population used by Preston, Heuveline, and Guillot (2001: 138–141), so the reader can refer to that book to find additional explanation. Suppose there is a hypothetical population in which all individuals die before age 5, defined by the following life table:

**Table 1: Life table (survival schedule) used in the data simulation**

| $a$      | 0    | 1    | 2    | 3    | 4    | 5    |
|----------|------|------|------|------|------|------|
| $\ell_a$ | 1.00 | 0.60 | 0.40 | 0.20 | 0.05 | 0.00 |

Function  $\ell_a$  gives the probability of surviving from age  $[0, 1)$  to age  $[a, a + 1)$ . All cohorts are driven by this survival schedule. We start with an initial population of  $N_0(t - 5) = 10^5$  individuals at age  $[0, 1)$  at time  $t - 5$ . At time  $t - 4$ ,  $N_1(t - 4) = N_0(t - 5) \ell_1$  individuals will reach age  $[1, 2)$ , and  $N_0(t - 4) = N_0(t - 5) e^r$  newly born will join the population, since  $r$  is the constant growth rate. Following the same procedure for older ages and assuming a growth rate of  $r = 0.01$ , at time  $t$  the population becomes stable with a population size of 234,248 individuals. The population counts by age are displayed in Table 2.

**Table 2: Population counts by age at time  $t$  of a simulated stable population with constant growth rate  $r = 0.01$**

| $a$   | $N_a(t)$   |
|-------|--|
| 0     | $N_0(t - 5) e^{5r} \ell_0 = 10^5 e^{0.05} 1.00 \approx 105, 127$ |
| 1     | $N_0(t - 5) e^{4r} \ell_1 = 10^5 e^{0.04} 0.60 \approx 62, 449$  |
| 2     | $N_0(t - 5) e^{3r} \ell_2 = 10^5 e^{0.03} 0.40 \approx 41, 218$  |
| 3     | $N_0(t - 5) e^{2r} \ell_3 = 10^5 e^{0.02} 0.20 \approx 20, 404$  |
| 4     | $N_0(t - 5) e^r \ell_4 = 10^5 e^{0.01} 0.05 \approx 5, 050$      |
| 5     | $N_0(t - 5) \ell_5 = 10^5 0.00 = 0$                              |
| Total | $N(t) = 234, 248$ individuals                                    |

From this population of 234,248 individuals we draw 100 random samples of 1,000 individuals each. Then, ignoring the ages, we follow all of these samples until extinction, recording at each time step the number of individuals who are still alive ( $\tilde{N}_a(t)$  values).

At each iteration, individuals randomly die depending on the corresponding age-specific probabilities of death. Applying (9\*), (10\*), and (11\*), we get estimates for the survival schedule and the stable age structure for all samples. Finally, combining the results from the 100 samples, we obtain mean values and empirical confidence intervals for all the estimates and compare them with the survival schedule used in the data simulation and the age structure of the whole population. The results displayed in Table 3 show that our method to estimate  $\ell_a$  and  $c_a$  for stable populations may provide serviceably accurate results. Differences among the estimates and the true values observed at the population level are due to the randomness in the data simulation and the sampling process, and also because of the rounding to integer values of the population counts. All true values fall within the 95% empirical confidence intervals of the estimates.

**Table 3: Estimates of the cohort survival schedule  $\hat{\ell}_a$  and the age structure  $\hat{c}_a$  of a simulated stable population, using information on life left**

| <i>a</i> | True values at the population level |         | Estimates from the samples |                            |
|----------|-------------------------------------|---------|----------------------------|----------------------------|
|          | $\ell_a$                            | $c_a$   | $\hat{\ell}_a$ (10*)       | $\hat{c}_a$ (11*)          |
| 0        | 1.00                                | 0.44879 | 1.00000                    | 0.44847 [0.42200, 0.47662] |
| 1        | 0.60                                | 0.26659 | 0.60131 [0.52142, 0.68716] | 0.26636 [0.24147, 0.28867] |
| 2        | 0.40                                | 0.17596 | 0.40345 [0.33590, 0.46453] | 0.17700 [0.15225, 0.19604] |
| 3        | 0.20                                | 0.08710 | 0.20128 [0.15901, 0.25306] | 0.08738 [0.07130, 0.10583] |
| 4        | 0.05                                | 0.02156 | 0.04807 [0.03019, 0.07151] | 0.02066 [0.01295, 0.02978] |
| 5        | 0.00                                | 0.00000 | 0.00000                    | 0.00000                    |

Note: Estimates from 100 samples of size 1,000 that are randomly taken from a stable population of 234,248 individuals. The samples are followed until death. Values between square brackets represent 95% empirical confidence intervals.

## 5. The 1805 cohort life table of Swedish females

As an example of application with actual data, we aim to recover the 1805 cohort life table of Swedish females by applying Equations (9\*) and (10\*) to data on population counts. The goal is to create a scenario in which we simulate the ‘capture’ of the Swedish females in 1805, ignoring their ages, and follow this group of individuals until extinction.

First, we ‘capture’ the population in 1805, as in a census. Next, we follow those individuals along the 19<sup>th</sup> century by recording the number of individuals still alive at each subsequent five-year period, obtaining the  $\tilde{N}_a(t)$  values. Finally, applying (9\*) we use these estimates to compute the number of individuals in each age group at each time period ( $N_a(t + a)$  values, see Figure 2), in order to apply (10\*) and estimate the survival schedule  $\ell_a$  of the 1805 cohort of Swedish females (Figure 3, left panel). All the

data comes from the Human Mortality Database (2018) (henceforth HMD). The results presented below are fully reproducible from the R code (R Core Team 2018) and data available in the supplementary materials.

## 5.1 Estimation of the growth rate

An estimate of the stable growth rate  $r$  is needed to apply the formulas in this article. This parameter can be estimated in various ways because all scalable stable population parameters grow exponentially at the same rate, including the number of births and deaths, and population size. Keyfitz and Caswell (2005, Chapter 5.2) provide an interesting discussion about how to estimate the growth rate from a single census, given the assumption of a stable population in which all cohorts follow the same life table. However, their approach requires knowledge of the number of individuals at a minimum of two different ages, and the probability of surviving between them. This is not the case in a context in which the ages of the population are unknown, and the survival schedule is estimated depending on the growth rate  $r$ , as shown in (2\*) and (10\*).

If the number of births  $B(t)$  at first observation can be assessed, one could simply consider

$$r = \frac{B(t) - D(t)}{N(t)} = \frac{B(t) - \tilde{D}(0, t)}{\tilde{N}(0, t)}.$$

When that information is not available,  $r$  can be estimated if the entire population (or a random sample) is observed at two time points  $t_1$  and  $t_2$  so that information is available on  $N(t_1)$  and  $N(t_2)$ . This may be the most convenient approach in many real-world scenarios. If that is the case,

$$(12) \quad r = \frac{\ln [N(t_2)] - \ln [N(t_1)]}{t_2 - t_1},$$

a formula that follows from the definition of a stable population (Keyfitz and Caswell 2005: 12). We use (12) to estimate the growth rate of the Swedish female population in the 19<sup>th</sup> century, using the total populations counts of 1805 and 1895 from the HMD:

$$r = \frac{\ln [N(1895)] - \ln [N(1805)]}{1895 - 1805} = \frac{\ln [2,509,076] - \ln [1,249,946]}{1895 - 1805} \approx 0.00774.$$

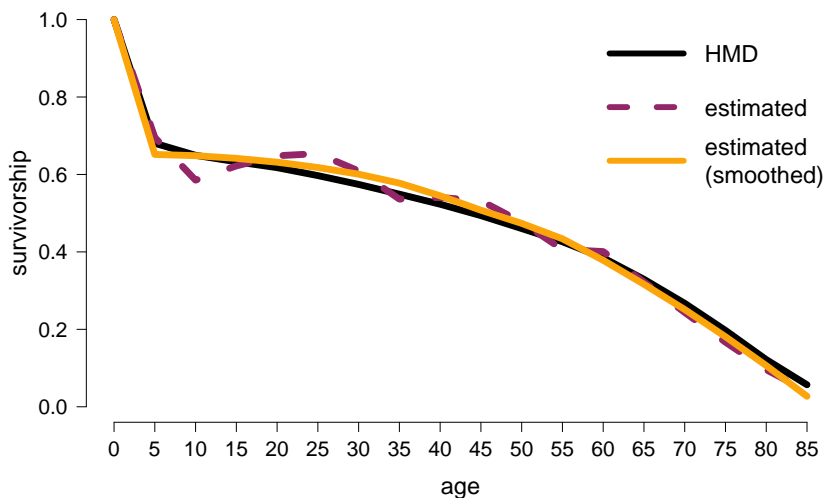
## 5.2 Reconstruction of the survival curve

Given the estimate of the population growth, we can proceed as in the example of Section 4.2, but with actual population counts from the HMD. First, we get the total population in 1805,  $N(1805) = \tilde{N}_0(1805) = 1,249,946$ . Next, we specify the number

of individuals from that initial population who are still alive in 1810. In terms of the available data, this corresponds to the number of Swedish females above age 5 in 1810,  $N_{5+}(1810) = \tilde{N}_5(1805) = 1,105,919$ . We follow the same procedure to obtain  $N_{10+}(1815) = \tilde{N}_{10}(1805) = 1,007,581$ ,  $N_{15+}(1820) = \tilde{N}_{15}(1805)$ , and so on. All these population counts were affected by migration flows. Nevertheless, in the stable population model, we assume closure to migration, something necessary to apply our method to estimate age-specific survival in a population with unknown ages.

Once the values  $\tilde{N}_0(1805), \dots, \tilde{N}_{90}(1805)$  are available, we can apply (9\*) and (10\*) to estimate the survival schedule of the 1805 cohort of Swedish females. Figure 4 presents the results by comparing the estimated survivorship with the survival curve from the corresponding cohort life table of the HMD.

**Figure 4:** Observed (HMD) and estimated survival schedule of the 1805 cohort of Swedish females using Equation (10\*)



Source: Human Mortality Database (2018).

The dark line in Figure 4 refers to the HMD life table. The dashed line is the result of applying (10\*) to the  $\tilde{N}_a(t)$  values. This curve is not monotonically decreasing, which contradicts the definition of the survival function. This is due to the fact that no population is strictly stable, and the Swedish population in the 19<sup>th</sup> century may have been affected by migration flows or by epidemics that may have increased mortality. To address this issue, the yellow line in Figure 4 shows the estimated survival curve smoothed using the R function `loess()` (R Core Team 2018). This function, in combination with



`predict()`, is one of the most common methods to smooth volatile data, and it is based on implementing a nonparametric approach that fits multiple regressions in a local neighborhood (see example of application in the R code of the supplementary materials). The similarity between the dark and yellow curves illustrates the validity of our theoretical approach to estimate age-specific survival in populations with unknown ages, but in which individuals are followed until death. Moreover, it also shows that the Swedish female population in the 19<sup>th</sup> century was rather stable.

## 6. Discussion

In this article, we have extended the concept of symmetries between life lived and left from stationary to stable populations. Equations (1\*) and (9\*) are a generalization to stable populations of previous results by Brouard (1989), Vaupel (2009), and Villavicencio and Riffe (2016), who limited their analysis to stationary populations. The stable population model represents a less restrictive setup with more potential areas of application. Further, by using the survival curve in (2\*), all the usual life table functions and statistics can be calculated, including the stable age structure.

The relationship between life lived and left may be especially relevant for the study of populations in which ages are unknown, but individuals are followed until death. In Section 4 we discuss some of the challenges of applying theoretical relationships to empirical data, especially due to the fact that in real-world scenarios time is not continuous, and some adjustments are necessary to move from a continuous to a discrete-time framework. In this regard, Equations (9\*), (10\*), and (11\*) may be more interesting to researchers dealing with actual data. Furthermore, it is remarkable that, contrary to what usually happens, the discrete equations are simpler to derive than the continuous ones, as illustrated in Figures 2 and 3. Their validity has been demonstrated in the applications of Sections 4.2 and 5.

It is worth mentioning some potential areas of application to human data. Historical records, for instance, usually present problems of age misreporting and of under-registration of births. If a census is carried out at some point in time and individuals are followed until death, that information could be used to estimate the age structure at the time of the census. Anthropologists could also employ a similar approach to study populations with lack of age information. Biologists may find these methods useful to estimate the survival and age structure of groups of individuals of unknown age in the wild that are followed over time. Finally, the approach suggested in this article could also be extended to multistate populations in which individuals transition from and to different states. For example, in health studies one could use the time until recovery from an illness as an input to estimate the distribution of the times of incidence until detection, if the assumption of stability is acceptable.

## References

- Brouard, N. (1986). Structure et dynamique des populations: La pyramide des années à vivre, aspects nationaux et exemples régionaux. *Espace, populations, sociétés* 4(2): 157–168. doi:10.3406/espos.1986.1120.
- Brouard, N. (1989). *Mouvements et modèles de population*. Yaoundé: Institut de Formation et de Recherche Démographiques. [http://iford-cm.org/images/documents\\_pedagogique/Mouvements-et-modeles-IFORD-1989.pdf](http://iford-cm.org/images/documents_pedagogique/Mouvements-et-modeles-IFORD-1989.pdf) (retrieved on June 24, 2018).
- Euler, L. (1970). A general investigation into the mortality and multiplication of the human species [translated by N. Keyfitz and B. Keyfitz]. *Theoretical Population Biology* 1(3): 307–314. doi:10.1016/0040-5809(70)90048-1.
- Goldstein, J.R. (2009). Life lived equals life left in stationary populations. *Demographic Research* 20(2): 3–6. doi:10.4054/DemRes.2009.20.2.
- Human Mortality Database (2018). [electronic resource]. Berkeley and Rostock: University of California and Max Planck Institute for Demographic Research. <http://www.mortality.org> (downloaded on June 4, 2018).
- Keyfitz, N. and Caswell, H. (2005). *Applied mathematical demography*. New York: Springer. doi:10.1007/b139042.
- Kim, Y.J. and Aron, J.L. (1989). On the equality of average age and average expectation of remaining life in a stationary population. *SIAM Review* 31(1): 110–113. doi:10.1137/1031005.
- Lotka, A.J. (1907). Relation between birth rates and death rates. *Science* 26(653): 21–22. doi:10.1126/science.26.653.21-a.
- Lotka, A.J. (1934). *Théorie analytique des associations biologiques: I. Principes*. Paris: Hermann.
- Lotka, A.J. (1939). *Théorie analytique des associations biologiques: II. Analyse démographique avec application particulière à l'espèce humaine*. Paris: Hermann.
- Lotka, A.J. (1998). *Analytical theory of biological populations* [translated by D.P. Smith and H. Rossert]. New York: Springer. doi:10.1007/978-1-4757-9176-1.
- Müller, H.G., Wang, J.L., Carey, J.R., Caswell-Chen, E.P., Chen, C., Papadopoulos, N., and Yao, F. (2004). Demographic window to aging in the wild: Constructing life tables and estimating survival functions from marked individuals of unknown age. *Aging Cell* 3(3): 125–131. doi:10.1111/j.1474-9728.2004.00096.x.
- Müller, H.G., Wang, J.L., Yu, W., Delaigle, A., and Carey, J.R. (2007). Survival and aging in the wild via residual demography. *Theoretical Population Biology* 72(4): 513–522.

[doi:10.1016/j.tpb.2007.07.003](https://doi.org/10.1016/j.tpb.2007.07.003).

- Preston, S.H., Heuveline, P., and Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Oxford: Blackwell.
- R Core Team (2018). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <http://www.r-project.org>.
- Riffe, T. (2015). The force of mortality by years lived is the force of increment by years left in stationary populations. *Demographic Research* 32(29): 827–834. [doi:10.4054/DemRes.2015.32.29](https://doi.org/10.4054/DemRes.2015.32.29).
- United Nations (1968). *The concept of a stable population: Application to the study of populations of countries with incomplete demographic statistics*. New York: United Nations, Department of Economic and Social Affairs, Population Division. <http://www.un.org/en/development/desa/population/publications/manual/model/stable-population.shtml> (retrieved on December 28, 2017).
- Vaupel, J.W. (2009). Life lived and left: Carey’s equality. *Demographic Research* 20(3): 7–10. [doi:10.4054/DemRes.2009.20.3](https://doi.org/10.4054/DemRes.2009.20.3).
- Villavicencio, F. and Riffe, T. (2016). Symmetries between life lived and left in finite stationary populations. *Demographic Research* 35(14): 381–398. [doi:10.4054/DemRes.2016.35.14](https://doi.org/10.4054/DemRes.2016.35.14).

