

# Chapter 4

## Sensitivity Analysis of Longevity and Life Disparity



### 4.1 Introduction

The population growth rate ( $\lambda$  or  $r$ ) analyzed in Chap. 3 is a population-level consequence of the individual-level vital rates. A similarly basic outcome, at the individual or cohort level, is longevity: the length of individual life. The most commonly encountered description of longevity is its expectation, the life expectancy. However, longevity is a random variable, differing among individuals (even when those individuals are subject to the same rates and hazards) because of the random vagaries of mortality and survival. Therefore, it is important to also consider its variance and higher moments. This chapter introduces the sensitivity analysis of longevity, which will be explored in more detail in Chaps. 5, 11, and 12.

As in Chap. 3, we will begin by reviewing a classic formula for the sensitivity of life expectancy in age-classified models. Then we will use matrix calculus to derive more general formulas for the moments of longevity, the distribution of age or stage at death, and the life disparity, applicable to age- or stage-classified populations.

### 4.2 Life Expectancy in Age-Classified Populations

**Notation** It is customary to denote life expectancy by symbols like  $e_x^0$  or  $e(x)$ , but in general the symbol  $e$  plays too many roles in mathematics to be helpful for our purposes. So, when we make the transition to matrix formulations, I will use the symbol  $\eta$ , in various vector and scalar manifestations, to indicate longevity.

Perturbation analysis of longevity has been pursued mostly within the framework of age-classified life cycles (e.g., Canudas Romo 2003; Keyfitz 1971; Pollard 1982; Vaupel 1986; Vaupel and Canudas Romo 2003). The life expectancy at age  $x$  is given by

$$e(x) = \frac{1}{\ell(x)} \int_x^\infty \ell(s) ds \quad (4.1)$$

where the survivorship function  $\ell(x)$  is the probability of survival to age  $x$ .

The classical result for the sensitivity of life expectancy at birth to a change in mortality at age  $a$  is

$$\frac{de(0)}{d\mu(a)} = -\ell(a)e(a). \quad (4.2)$$

That is, the sensitivity of life expectancy at birth to a change in mortality at age  $a$  is equal to the product of the probability of survival to age  $a$  and the life expectancy at age  $a$ . In other words,  $e(0)$  is most sensitive to changes in mortality at ages to which lots of individuals survive (to experience the change in mortality) and beyond which there is lots of longevity remaining (so they can enjoy the change in mortality). The derivative is negative because increasing mortality reduces life expectancy.

The result was presented independently by Keyfitz (1971) who also referenced some earlier approaches (Wilson 1938; Irwin 1949) and by Pollard (1982). Keyfitz's derivation was sketchy, and Pollard simply stated that the result was well-known, and gave no derivation. From a general sensitivity analysis perspective, we can derive the result using the same approach applied in Chap. 3 to population growth rate.

### 4.2.1 Derivation

Differentiating (4.1) with respect to mortality at some specified age  $a$  gives

$$\frac{de(0)}{d\mu(a)} = \int_0^\infty \frac{d\ell(s)}{d\mu(a)} ds \quad (4.3)$$

and our problem reduces to finding the derivative of  $\ell(s)$  with respect to  $\mu(a)$ . To do so, introduce a parameter  $\theta$  to measure the size of the perturbation at age  $a$ , and write mortality as

$$\mu(x, \theta) = \mu(x, 0) + \theta \delta(x - a) \quad (4.4)$$

where  $\delta(x - a)$  is the Dirac delta function.<sup>1</sup> The derivative with respect to  $\mu(a)$  is obtained by differentiating with respect to  $\theta$  and evaluating the result at  $\theta = 0$ .

---

<sup>1</sup>See Chap. 3 for a description of this generalized function.

Write survivorship as

$$\ell(x, \theta) = \exp \left[ - \int_0^x \mu(z, \theta) dz \right] \quad (4.5)$$

so that

$$\frac{d\ell(x, \theta)}{d\theta} = -\ell(x, \theta) \int_0^x \frac{d\mu(z, \theta)}{d\theta} dz \quad (4.6)$$

From (4.4) we have

$$\frac{d\mu(z, \theta)}{d\theta} = \delta(z - a) \quad (4.7)$$

so that

$$\frac{d\ell(x, \theta)}{d\theta} = -\ell(x, \theta) \int_0^\infty \delta(z - a) dz \quad (4.8)$$

$$= -\ell(x, \theta) H(x - a) \quad (4.9)$$

where  $H(\cdot)$  is the unit step function. Substituting this into (4.3) and evaluating at  $\theta = 0$  gives

$$\frac{de(0)}{d\mu(a)} = - \int_0^\infty \ell(s) H(s - a) ds \quad (4.10)$$

$$= - \int_a^\infty \ell(s) ds \quad (4.11)$$

which, by (4.1) is equal to (4.2).

### 4.3 A Markov Chain Model for the Life Cycle

Age has a special status in demography because it is continuous, linear, and permits movement in only one direction and at one rate (age increases by one unit for every unit of time). All other demographic characteristics have the potential for much greater flexibility, and the operators that describe movement and development of individuals require an equal degree of flexibility. This book is devoted to matrix formulations of these problems, which have the great advantage of permitting both age and stage-classified models. The basic formulation, as far as longevity is concerned, is that of a finite-state absorbing Markov chain.

### 4.3.1 *A Markov Chain Formulation of the Life Cycle*

We describe the life cycle as an absorbing Markov chain. This approach was pioneered in demography by Feichtinger (1971) and Hoem (1969), and has been greatly extended in recent years (Caswell 2001, 2006, 2009; Horvitz and Tuljapurkar 2008; Tuljapurkar and Horvitz 2006; Steinsaltz and Evans 2004). Good sources for the basic theory of absorbing Markov chains are Kemeny and Snell (1976) and Iosifescu (1980).

These models will be explored in more detail in Chaps. 5 and 11. The sensitivity analysis of measures of variance in longevity has been developed by Van Raalte and Caswell (2013) and Engelman et al. (2014). An important extension of Markov chain models for longevity is the incorporation of “rewards” to represent the value, in some sense, of the length of life, extending methods developed for dynamic programming (Howard 1960). The rewards include the production of offspring (Caswell 2011; van Daalen and Caswell 2015, 2017), the accumulation of income and expenditures (Caswell and Kluge 2015) and healthy longevity (Caswell and Zarulli 2018). The sensitivity analysis of these important models is derived in van Daalen and Caswell (2017).

Markov chain theory distinguishes between *recurrent* and *transient* states. A recurrent state has the property that the probability of returning to that state at least once is 1. A transient state is one for which that probability is less than 1. If a Markov chain contains transient states, it will eventually leave those states and arrive in a recurrent state or class of states, where it will remain permanently. Such a chain is called *absorbing*. Absorbing chains are the basic model for the demography of individuals because life is inherently transient. Any individual will, with probability one, eventually leave the set of living states and be absorbed by death.

If a Markov chain consists of a single set of recurrent states that all communicate with each other, it is said to be ergodic. The transition matrix for an ergodic chain is irreducible and primitive. Ergodic Markov chains play a limited role in demographic contexts because they cannot include mortality. Chapter 11 will, however, present the sensitivity analysis of these models.

In demographic models, individuals move among a set of transient (i.e., living) states in their life cycle before they eventually reach an absorbing state (death). Transient states may represent age classes, developmental or life history stages, or states defined by health, employment, economic, or other kinds of status. In studying longevity, we are particularly interested in absorbing states representing death, or perhaps death classified by age or stage at death, or by cause of death. The analysis applies equally to other ways of leaving the life cycle (e.g., graduation in a model of educational states, discharge from treatment in model of health states).

Number the stages in the life cycle so that the transient states are  $1, \dots, s$  and the absorbing states are  $s + 1, \dots, s + a$ . Then the transition matrix of the Markov chain is

$$\mathbf{P} = \left( \begin{array}{c|c} \mathbf{U} & \mathbf{0} \\ \hline \mathbf{M} & \mathbf{I} \end{array} \right) \tag{4.12}$$

Here,  $\mathbf{U}$  is the  $s \times s$  matrix of transition probabilities among the transient states. The  $a \times s$  matrix  $\mathbf{M}$  gives the probabilities of absorption in each of the absorbing states. The columns of  $\mathbf{P}$  sum to one. I assume that the spectral radius (the dominant eigenvalue) of  $\mathbf{U}$  is strictly less than one; a sufficient condition for this is that there is a non-zero probability of ultimate death for every stage.

Age-classified models are a special case with survival probabilities on the subdiagonal (and possibly in the last diagonal entry); e.g., for  $s = 3$  in which

$$\mathbf{U} = \begin{pmatrix} 0 & 0 & 0 \\ p_1 & 0 & 0 \\ 0 & p_2 & p_3 \end{pmatrix} \tag{4.13}$$

The age-specific survival probability is  $p_i = e^{-\mu_i}$ , with  $\mu_i$  a mortality rate applying to age class  $i$ . The  $(s, s)$  entry of  $\mathbf{U}$  is an age-independent survival probability for a final open-ended age class, with a remaining life expectancy of  $1/(1-p_s)$ . If  $p_s = 0$  no one survives beyond age class  $s$ . When the age-classified model is constructed from a life table,  $p_i = 1 - q_{i-1}$ ; that is, the survival of age-class 1 is the complement of the probability of death between age 0 and 1.

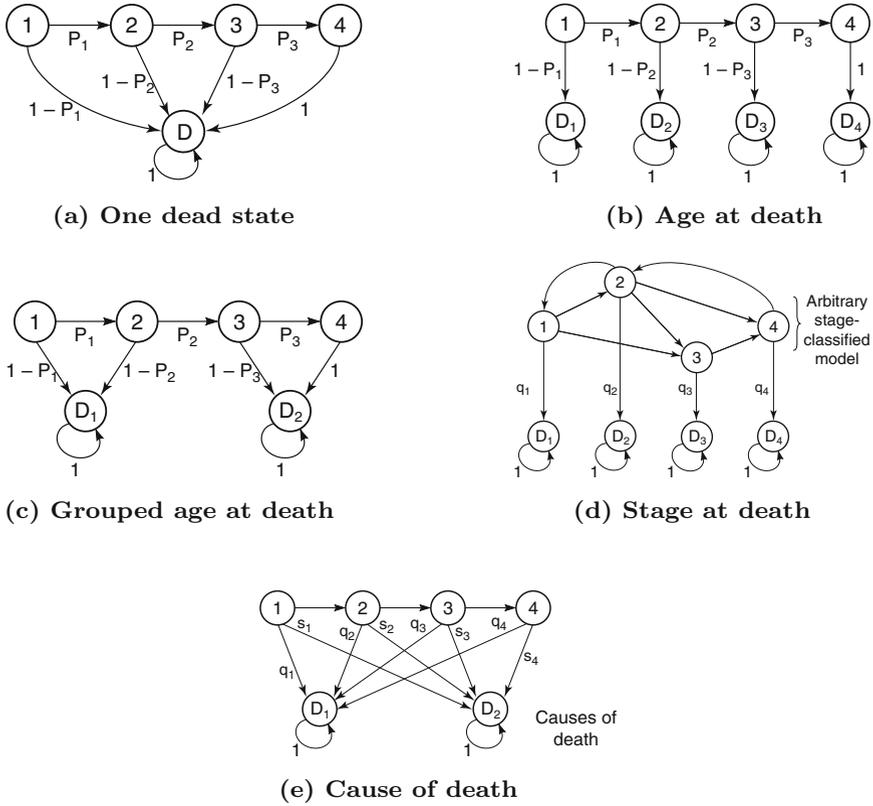
The mortality matrix  $\mathbf{M}$  gives the probabilities of transition from each of the transient states to each of the absorbing states. Figure 4.1 shows some examples of life cycle formulations that can arise, including both age and stage classification in the transient states, and absorbing states classified by age at death, grouped ages at death, stage at death, or cause of death. The resulting mortality matrices are

Figure 4.1a  $\mathbf{M} = (1 - P_1 \ 1 - P_2 \ 1 - P_3 \ 1)$  (4.14)

Figure 4.1b  $\mathbf{M} = \begin{pmatrix} 1 - P_1 & 0 & 0 & 0 \\ 0 & 1 - P_2 & 0 & 0 \\ 0 & 0 & 1 - P_3 & 0 \\ 0 & 0 & 0 & 1 - P_4 \end{pmatrix}$  (4.15)

Figure 4.1c  $\mathbf{M} = \begin{pmatrix} 1 - P_1 & 1 - P_2 & 0 & 0 \\ 0 & 0 & 1 - P_3 & 1 - P_4 \end{pmatrix}$  (4.16)

Figure 4.1d  $\mathbf{M} = \begin{pmatrix} q_1 & 0 & 0 & 0 \\ 0 & q_2 & 0 & 0 \\ 0 & 0 & q_3 & 0 \\ 0 & 0 & 0 & q_4 \end{pmatrix}$  (4.17)



**Fig. 4.1** Life cycle graphs showing some alternative choices for structure of the absorbing state: death, age at death, stage at death, or cause of death. **(a)** Age-classified with one dead state. **(b)** Age-classified, age at death. **(c)** Age-classified, grouped ages at death. **(d)** Stage-classified, stage at death. **(e)** Age-classified, causes of death

Figure 4.1e 
$$\mathbf{M} = \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ s_1 & s_2 & s_3 & s_4 \end{pmatrix} \tag{4.18}$$

The beauty of formulating longevity as a Markov chain is that many statistics of longevity can be written in terms of the matrices  $\mathbf{U}$  and  $\mathbf{M}$  and sensitivity analysis can be carried out using matrix calculus.

### 4.3.2 Occupancy Times

Consider an individual in transient state  $j$ . Eventual absorption is certain. But before that, the individual will occupy various transient states. The number of such visits,

the occupancy time<sup>2</sup> is the basic unit of longevity. Occupancy is particularly central in studies of health demography, where it quantifies the parts of a life spent in different health states. But, even without the added dimension of something like health, occupancy of transient states is the basis of longevity analysis.

Let  $v_{ij}$  be the number of visits to transient state  $i$  by an individual in transient state  $j$ , prior to absorption. Its expectation is given by the fundamental matrix (e.g., Kemeny and Snell 1976; Iosifescu 1980)

$$\mathbf{N} = (E(v_{ij})) \quad (4.19)$$

$$= (\mathbf{I} - \mathbf{U})^{-1} \quad (4.20)$$

More details, and examples, for the higher moments and variances of occupancy times are given in Chaps. 5 and 11.

### 4.3.3 Longevity

The longevity of an individual in state  $j$  can be equated to the total occupancy time of all transient states by that individual, prior to eventual absorption. Let  $\eta_j$  be this longevity; the expectation of  $\eta_j$  is the sum of the elements in column  $j$  of  $\mathbf{N}$ . We define  $\boldsymbol{\eta}_1$  and  $\boldsymbol{\eta}_2$  as the vectors containing the first and second moments of longevity, respectively. Then

$$E(\boldsymbol{\eta})^T = \boldsymbol{\eta}_1^T = \mathbf{1}^T \mathbf{N} \quad (4.21)$$

Figure 4.2a shows the life expectancy for India in 1961 and Japan in 2006.

The vector of the second moments of longevity satisfies

$$\boldsymbol{\eta}_2^T = \boldsymbol{\eta}_1^T (2\mathbf{N} - \mathbf{I}) \quad (4.22)$$

(Iosifescu 1980). The variance and standard deviation of longevity are thus

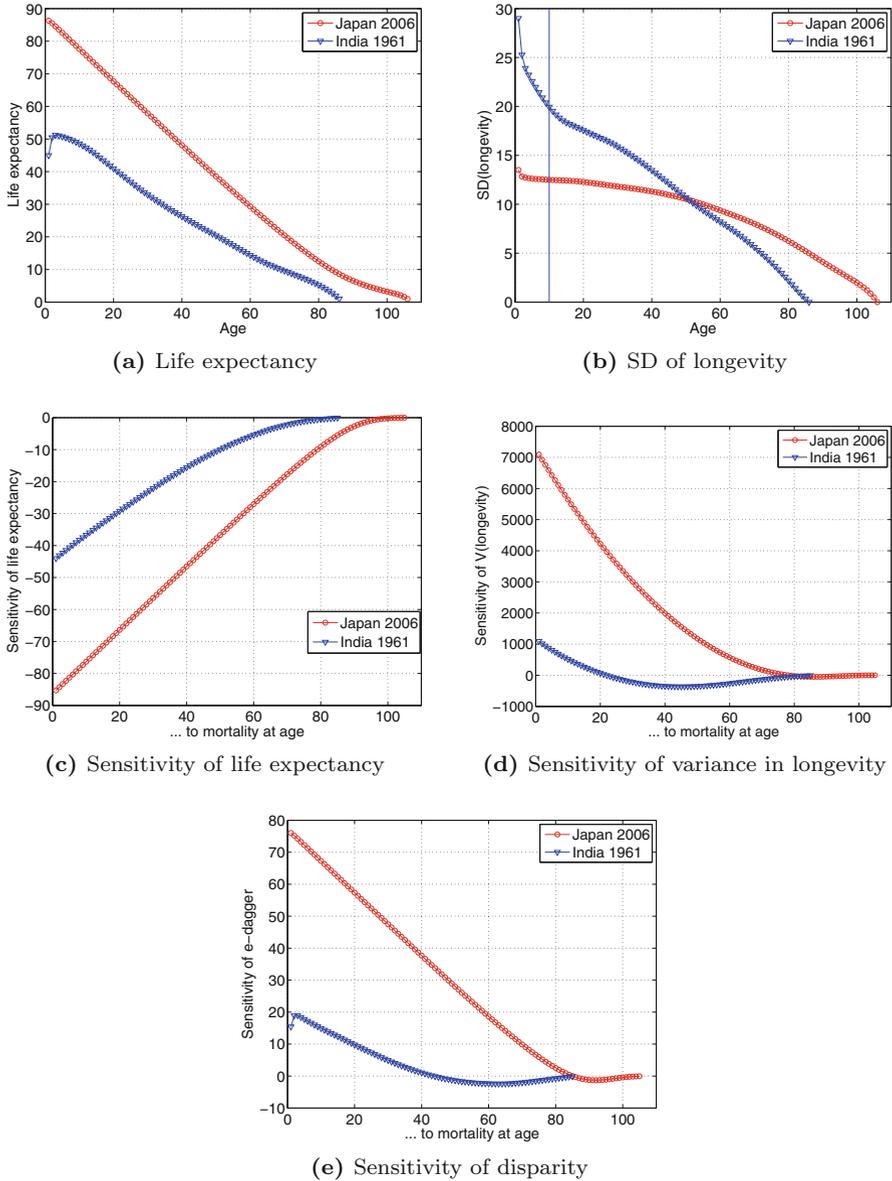
$$V(\boldsymbol{\eta})^T = \boldsymbol{\eta}_2 - \boldsymbol{\eta}_1 \circ \boldsymbol{\eta}_1 \quad (4.23)$$

$$SD(\boldsymbol{\eta}) = \sqrt{V(\boldsymbol{\eta})} \quad (4.24)$$

where the square root is taken element-wise.

---

<sup>2</sup>Because time is discrete here, the number of visits is equal to the number of time increments, which is the amount of time spent in the state. In continuous-time models, the number of visits to, and the length of time spent in, a transient state are different. The corresponding calculations for continuous-time models are given in Chap. 12.



**Fig. 4.2** Calculations for longevity of India (1961) and Japan (2006). (a) Remaining life expectancy as a function of age. (b) Standard deviation of remaining longevity as a function of age. Vertical line at age 10 indicates  $SD_{10}$ , sometimes used as a measure of lifespan disparity. (c) Sensitivity of life expectancy at birth to changes mortality at each age. (d) Sensitivity of variance in longevity at birth to changes in mortality at each age. (e) Sensitivity of life disparity  $\eta^\dagger$  to changes in mortality at each age

Note that  $V(\boldsymbol{\eta})$  and  $SD(\boldsymbol{\eta})$  are vectors; their elements give the variance or standard deviation of longevity for individuals in each stage, making it easy to examine variation in remaining longevity conditional on the starting age. This conditioning can be important; Edwards and Tuljapurkar (2005) have made a strong case that  $SD(\eta_{10})$ , starting from age 10, is a good index to prevent infant and child mortality from obscuring patterns in old age longevity.

Figure 4.2b shows  $SD(\boldsymbol{\eta})$  for India and Japan. The standard deviation at birth,  $SD(\eta_1)$  is roughly twice as great in India as in Japan, a discrepancy that remains at  $SD(\eta_{10})$ . Eventually, beyond the age of 50,  $SD(\boldsymbol{\eta})$  becomes greater in India than in Japan.

#### 4.3.4 Age or Stage at Death

If the model contains more than one absorbing state (as in all the cases but the first in Fig. 4.1), the eventual fate of an individual is uncertain. The probability distributions of the eventual absorbing state are given by the columns of the matrix

$$\mathbf{B} = \mathbf{M}\mathbf{N} \quad (4.25)$$

where  $b_{ij}$  is the probability of eventual absorption in absorbing state  $i$  for an individual starting in transient state  $j$  (Iosifescu 1980).

Suppose that the absorbing stages are defined as the age (or stage) at death, as in Fig. 4.1b, d. Then  $\mathbf{M}$  is given by Eq. (4.17) and the  $j$ th column of  $\mathbf{B}$  is the probability distribution of age at death for an individual starting in age class  $j$ :

$$\boldsymbol{\psi}_j = \mathbf{B}(:, j) = \mathbf{B}\mathbf{e}_j. \quad (4.26)$$

#### 4.3.5 Life Lost and Life Disparity

When an individual dies, it loses the remaining life that it would have experienced, had it not died. This counterfactual proposition seems abstract, but we can make it concrete by asking for the expectation of that lost lifetime. An individual that dies at age  $x$  will lose, on average, an amount of life given by the life expectancy at age  $x$ . Averaging this remaining life expectancy over the distribution of age at death gives the mean life lost due to mortality. Vaupel and Canudas Romo (2003) denoted the life lost by  $e^\dagger$ . Here we define the vector  $\boldsymbol{\eta}^\dagger$ , whose  $i$ th entry is the expected life lost due to mortality by an individual starting in age class  $i$ ; it is given by

$$\left(\boldsymbol{\eta}^\dagger\right)^T = \boldsymbol{\eta}_1^T \mathbf{B}. \quad (4.27)$$

Calculations of life lost from mortality due to specific causes of death play a central role in the calculations of disability-adjusted life years (DALYs) used in calculations of the burden of diseases (e.g., Devleeschauwer et al. 2014; GBD 2016 DALYs and HALE Collaborators 2017). See Caswell and Zarulli (2018) for the relationship between DALY calculations and Markov chain methods, and for a calculation of the variance in life lost.

The life lost  $\eta^\dagger$  has an additional interpretation as a measure of disparity. Consider a population in which everyone dies at the same age. In such a situation,  $\eta^\dagger = 0$ , because at the age of death, there is no additional life expectancy. Thus  $\eta^\dagger$  is a measure of “life disparity;” the larger its value, the more disparity there is among individuals in age at death (Vaupel et al. 2011).

The values of life disparity in age class 1, for Japan and India, in years, are

$$\eta_1^\dagger = \begin{cases} 10.1 & \text{Japan} \\ 23.9 & \text{India} \end{cases} \quad (4.28)$$

Just as India has a much larger variance in longevity than Japan, it also has a higher life disparity.

## 4.4 Sensitivity Analysis

Our goal is to obtain expressions for the derivatives of  $E(\eta)$ ,  $V(\eta)$ ,  $SD(\eta)$ ,  $\mathbf{B}$ , and  $\eta^\dagger$ , with respect to changes in age specific-mortality rates. The calculations and some results (contrasting the mortality schedules of Japan and India) are given here. More details are presented in Chaps. 5 and 11. Results are presented in terms of an arbitrary vector  $\theta$  of parameters on which  $\mathbf{U}$  and  $\mathbf{M}$  depend. In the examples,  $\theta$  will be the vector  $\mu$  of age-specific mortality rates.

### 4.4.1 Sensitivity of the Fundamental Matrix

The fundamental matrix  $\mathbf{N}$  appears in many of these formulas. Its sensitivity was first obtained by Caswell (2006). Suppose that  $\mathbf{U}$  is a function of some vector  $\theta$  of parameters. Then

$$\frac{d\text{vec } \mathbf{N}}{d\theta^\top} = (\mathbf{N}^\top \otimes \mathbf{N}) \frac{d\text{vec } \mathbf{U}}{d\theta^\top} \quad (4.29)$$

(see Chap. 5).

### 4.4.2 Sensitivity of Life Expectancy

The sensitivity of the vector of life expectancy as a function of age is obtained by differentiating (4.21),

$$d\eta_1^\top = \mathbf{1}^\top(d\mathbf{N}) \quad (4.30)$$

Applying the vec operator and Roth's theorem (2.13) gives

$$d\eta_1 = \left( \mathbf{I} \otimes \mathbf{1}^\top \right) d\text{vec } \mathbf{N} \quad (4.31)$$

$$= \left( \mathbf{I} \otimes \mathbf{1}^\top \right) \left( \mathbf{N}^\top \otimes \mathbf{N} \right) d\text{vec } \mathbf{U} \quad (4.32)$$

$$= \left( \mathbf{N}^\top \otimes \eta_1^\top \right) d\text{vec } \mathbf{U}. \quad (4.33)$$

The last step uses the fact that  $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$ . Applying the chain rule and the first identification theorem gives the result

$$\frac{d\eta_1}{d\theta^\top} = \left( \mathbf{N}^\top \otimes \eta_1^\top \right) \frac{d\text{vec } \mathbf{U}}{d\theta^\top} \quad (4.34)$$

**Sensitivity to mortality** If interest focuses on changes in age-specific mortality, so that  $\theta = \mu$ , then the sensitivity formula expands, using the chain rule, to

$$\frac{d\eta_1}{d\mu^\top} = \left( \mathbf{N}^\top \otimes \eta_1^\top \right) \frac{d\text{vec } \mathbf{U}}{d\mu^\top} \quad (4.35)$$

This can be evaluated in several ways, depending on how the matrix  $\mathbf{U}$  is written as a function of mortality. One approach is used in Sect. 4.4.3, and a somewhat more widely useful approach in Sect. 4.4.4.

The results for Japan and India are shown in Fig. 4.2. Life expectancy is more sensitive to changes in mortality in Japan than in India; the (absolute value of) sensitivity decreases almost linearly with age in Japan, and slightly less linearly in India (Fig. 4.2). On the other hand, life expectancy is more elastic to changes in mortality in India, and less so in Japan.

### 4.4.3 Generalizing the Keyfitz-Pollard Formula

The Keyfitz-Pollard formula for the sensitivity of life expectancy to changes in mortality rate, given in Eq. (4.2), has a clear interpretation: the sensitivity to mortality at age  $a$  depends on the probability of survival to age  $a$  and the remaining life expectancy at age  $a$ . We are now in a position to generalize this to stage-classified matrix models.

First, we derive the matrix version of the Keyfitz-Pollard result, for the sensitivity of life expectancy of age class 1, which is

$$dE(\eta_1) = (\mathbf{e}_1^\top \otimes \mathbf{1}^\top) d\text{vec } \mathbf{N} \quad (4.36)$$

$$= (\mathbf{e}_1^\top \otimes \mathbf{1}^\top) (\mathbf{N}^\top \otimes \mathbf{N}) d\text{vec } \mathbf{U} \quad (4.37)$$

Consider a population with  $s$  age classes and let  $\mu_i$  be the mortality rate and  $p_i = \exp(-\mu_i)$  the survival probability for age class  $i$ . The matrix  $\mathbf{U}$  is given by (4.13), which can be written

$$\mathbf{U} = \sum_{k=1}^{s-1} (\mathbf{e}_{k+1} \mathbf{e}_k^\top) p_k \quad (4.38)$$

where  $\mathbf{e}_k$  is the unit vector, of length  $s$ , with a 1 in the  $k$ th position and zeros elsewhere. Differentiating  $\mathbf{U}$  and applying the vec operator gives

$$d\text{vec } \mathbf{U} = - \sum_{k=1}^{s-1} (\mathbf{e}_k \otimes \mathbf{e}_{k+1}) p_k (d\mu_k) \quad (4.39)$$

Substitute (4.39) into (4.37) and consider a perturbation of mortality at age  $a$ ; the result is

$$\frac{dE(\eta_1)}{d\mu_a} = - (\mathbf{e}_1^\top \otimes \mathbf{1}^\top) (\mathbf{N}^\top \otimes \mathbf{N}) (\mathbf{e}_a \otimes \mathbf{e}_{a+1}) p_a. \quad (4.40)$$

This simplifies to

$$\frac{dE(\eta_1)}{d\mu_a} = - (\mathbf{e}_1^\top \mathbf{N}^\top \mathbf{e}_a \otimes \mathbf{1}^\top \mathbf{N} \mathbf{e}_{a+1}) p_a \quad (4.41)$$

$$= - \underbrace{E(v_a)}_{\text{survival}} p_a \underbrace{E(\eta_{a+1})}_{\text{expectancy}} \quad \text{age-classified} \quad (4.42)$$

In an age-classified model,  $v_a$  is either 0 or 1 (you cannot occupy a year of age for more than 1 year); hence the  $E(v_a)$  is the probability of survival to age  $a$ . Thus we have a matrix version of the Keyfitz-Pollard result: the sensitivity of life expectancy is the probability of survival to age  $a$  times the probability of survival from  $a$  to  $a + 1$ , times the life expectancy at age  $a + 1$ .

Now apply the same approach to a stage-classified model, in which  $\mathbf{U}$  can be written as the product of a diagonal matrix  $\mathbf{\Sigma}$  with survival probabilities on the diagonal, and a stochastic matrix  $\mathbf{G}$  giving the transition probabilities conditional on survival:

$$\mathbf{U} = \mathbf{G}\boldsymbol{\Sigma} \quad (4.43)$$

$$= \mathbf{G} \begin{pmatrix} p_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_s \end{pmatrix} \quad (4.44)$$

$$= \mathbf{G} \sum_{k=1}^s (\mathbf{e}_k \mathbf{e}_k^T) p_k \quad (4.45)$$

Differentiating and applying the vec operator gives

$$d\text{vec } \mathbf{U} = \sum_{k=1}^s (\mathbf{e}_k \otimes \mathbf{G}\mathbf{e}_k) p_k (d\mu_k) \quad (4.46)$$

Substitute this into (4.37) and focus on a change in mortality at stage  $a$ ; the result is

$$\frac{dE(\eta_1)}{d\mu_a} = -(\mathbf{e}_1^T \otimes \mathbf{1}^T) (\mathbf{N}^T \otimes \mathbf{N}) (\mathbf{e}_a \otimes \mathbf{G}\mathbf{e}_a) p_a \quad (4.47)$$

which simplifies to

$$\frac{dE(\eta_1)}{d\mu_a} = -(\mathbf{e}_1^T \mathbf{N}^T \mathbf{e}_a \otimes \mathbf{1}^T \mathbf{N} \mathbf{G}\mathbf{e}_a) p_a \quad (4.48)$$

$$= -E(v_{a1}) E(\boldsymbol{\eta}^T) \mathbf{G}(:, a) p_a \quad (4.49)$$

$$= -\underbrace{E(v_{a1})}_{\text{occupancy}} \sum_{h=1}^s \underbrace{p_a g_{ha}}_{\text{transitions}} \underbrace{E(\eta_h)}_{\text{expectancy}} \quad \text{stage-classified} \quad (4.50)$$

Equation (4.50) is the stage-classified version of Keyfitz-Pollard: the sensitivity of life expectancy to a change in mortality in stage  $j$  is the product of the expected time spent in stage  $j$  and the remaining life expectancy, calculated as an average of the life expectancy of all stages  $k$ , weighted by the probability of transition from  $j$  to  $k$ . This can be simplified further by noting that, for either age or stage-classified populations,  $\mathbf{G}(:, a) p_a = \mathbf{U}(:, a)$ , so that a completely general expression is

$$\frac{dE(\eta_1)}{d\mu_a} = -E(v_{a1}) E(\boldsymbol{\eta}^T) \mathbf{U}(:, a) \quad \text{age- or stage-classified} \quad (4.51)$$

#### 4.4.4 Sensitivity of the Variance of Longevity

The sensitivity of the variance in longevity is obtained by differentiating (4.23)

$$dV(\eta) = d\eta_2 - 2(\eta_1 \circ d\eta_1) \quad (4.52)$$

and applying the vec operator (using results from Chap. 2 on the vec of the Hadamard product), to obtain

$$dV(\eta) = d\eta_2 - 2\mathcal{D}(\eta_1)d\eta_1. \quad (4.53)$$

The derivative of  $\eta_1$  is already given by (4.33):

$$d\eta_1 = (\mathbf{N}^\top \otimes \eta_1^\top) d\text{vec } \mathbf{U}. \quad (4.54)$$

The derivative of  $\eta_2$  is obtained by differentiating (4.22):

$$d\eta_2^\top = 2(d\eta_1^\top)\mathbf{N} + d\eta_1^\top(d\mathbf{N}) - d\eta_1^\top \quad (4.55)$$

Applying the vec operator to both sides and substituting (4.29) for  $d\text{vec } \mathbf{N}$  gives

$$d\eta_2 = (2\mathbf{N}^\top - \mathbf{I})d\eta_1 + 2(\mathbf{N}^\top \otimes \eta_1^\top \mathbf{N}) d\text{vec } \mathbf{U} \quad (4.56)$$

Inserting (4.54) for  $d\eta_1$  and (4.56) for  $d\eta_2$  into (4.53) gives the sensitivity of the variance in remaining longevity, for any starting age or stage, to changes in  $\mathbf{U}$ . The sensitivity of longevity to mortality is obtained by differentiating  $\mathbf{U}$  with respect to  $\boldsymbol{\mu}$ .

**Derivatives of  $\mathbf{U}$**  The derivative of  $\mathbf{U}$  to the mortality vector  $\boldsymbol{\mu}$  are obtained as follows. For an age-classified model, define an age-advancement matrix

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & [1] \end{pmatrix} \quad (4.57)$$

(show here for three age classes, with the optional open-ended last age class). This matrix will mask the entries of a matrix  $\mathbf{1}\mathbf{p}^\top$ , that contains  $\mathbf{p}$  in each row, to obtain

$$\mathbf{U} = \mathbf{L} \circ (\mathbf{1}\mathbf{p}^\top) \quad (4.58)$$

Differentiating and applying the vec operator gives

$$d\mathbf{U} = \mathbf{L} \circ (\mathbf{1}(d\mathbf{p}^\top)) \quad (4.59)$$

$$d\text{vec } \mathbf{U} = \mathcal{D}(\text{vec } \mathbf{L})(\mathbf{I} \otimes \mathbf{1}) d\mathbf{p}. \quad (4.60)$$

Since  $\mathbf{p} = \exp(-\boldsymbol{\mu})$ ,

$$d\mathbf{p} = -\mathcal{D}(\mathbf{p})d\boldsymbol{\mu}, \quad (4.61)$$

and hence

$$d\text{vec } \mathbf{U} = -\mathcal{D}(\text{vec } \mathbf{L}) (\mathbf{I} \otimes \mathbf{1}) \mathcal{D}(\mathbf{p})d\boldsymbol{\mu} \quad \text{age-classified} \quad (4.62)$$

For a stage-classified model, write  $\mathbf{U} = \mathbf{G}\boldsymbol{\Sigma}$ , as in (4.44) as

$$\mathbf{U} = \mathbf{G} \left[ \mathbf{I} \circ (\mathbf{1}\mathbf{p}^\top) \right] \quad (4.63)$$

Differentiating and applying the vec operator, following the strategy of (4.60), gives

$$d\text{vec } \mathbf{U} = -(\mathbf{I} \otimes \mathbf{G}) \mathcal{D}(\text{vec } \mathbf{I}) (\mathbf{I} \otimes \mathbf{1}) \mathcal{D}(\mathbf{p})d\boldsymbol{\mu} \quad \text{stage-classified} \quad (4.64)$$

Substituting (4.62) and (4.64) into the expressions for  $d\eta_1$  and  $d\eta_2$ , and substituting those into (4.53) gives the sensitivity of the variance in longevity to age- or stage-specific mortality. It is possible to carry out the substitutions and to arrive at a single (large) expression for  $dV(\boldsymbol{\eta})$ ; see Chap. 5.

Figure 4.2d shows the sensitivity and elasticity of variance of longevity to changes in age-specific mortality. The variance is more sensitive to mortality changes in Japan than in India, and the sensitivities are highest at young ages. Both life tables have the property that sensitivities are positive at early ages ( $\approx 0$ –20 for India,  $\approx 0$ –80 for Japan) and then become negative. Before this age, reductions in mortality will reduce variance; after this age, reductions in mortality increase the variance. See Sect. 4.4.6 for more on this.

#### 4.4.5 Sensitivity of the Distribution of Age at Death

The sensitivity of the distribution of age or stage at death is obtained by differentiating (4.25) and applying the vec operator,

$$d\text{vec } \mathbf{B} = (\mathbf{N}^\top \otimes \mathbf{I}) d\text{vec } \mathbf{M} + (\mathbf{I} \otimes \mathbf{M}) d\text{vec } \mathbf{N}. \quad (4.65)$$

We already know  $d\text{vec } \mathbf{N}$ . To obtain  $d\text{vec } \mathbf{M}$ , note that when the absorbing states are defined in terms of stage at death

$$\mathbf{M} = \mathbf{I} - \mathcal{D}(\mathbf{p}) \quad (4.66)$$

and thus

$$d\text{vec } \mathbf{M} = -\mathcal{D}(\text{vec } \mathbf{I})(\mathbf{I} \otimes \mathbf{1}) d\mathbf{p} \quad (4.67)$$

It is revealing to write the sensitivity of  $\mathbf{B}$  to changes in mortality using the chain rule,

$$\frac{d\text{vec } \mathbf{B}}{d\boldsymbol{\mu}^\top} = (\mathbf{N}^\top \otimes \mathbf{I}) \frac{d\text{vec } \mathbf{M}}{d\mathbf{p}^\top} \frac{d\mathbf{p}}{d\boldsymbol{\mu}^\top} + (\mathbf{I} \otimes \mathbf{M}) \frac{d\text{vec } \mathbf{N}}{d\text{vec }^\top \mathbf{U}} \frac{d\text{vec } \mathbf{U}}{d\mathbf{p}^\top} \frac{d\mathbf{p}}{d\boldsymbol{\mu}^\top} \quad (4.68)$$

and to recognize how many of the pieces we have already obtained.

The distribution of stage at death for individuals starting in stage  $j$  is given by column  $j$  of  $\mathbf{B}$ ; i.e.,  $\boldsymbol{\psi}_j = \mathbf{B}(:, j)$ . The sensitivity of  $\boldsymbol{\psi}_j$  to changes in mortality is

$$\frac{d\boldsymbol{\psi}_j}{d\boldsymbol{\mu}^\top} = (\mathbf{e}_j \otimes \mathbf{I}) \frac{d\text{vec } \mathbf{B}}{d\boldsymbol{\mu}^\top} \quad (4.69)$$

for any age or stage  $j$  of interest.

#### 4.4.6 Sensitivity of Life Disparity

To get the sensitivity of the vector  $\boldsymbol{\eta}^\dagger$ , differentiate and apply the  $\text{vec}$  operator to Eq. (4.27), which gives

$$d\boldsymbol{\eta}^\dagger = \mathbf{B}^\top d\boldsymbol{\eta}_1 + (\mathbf{I} \otimes \boldsymbol{\eta}_1^\top) d\text{vec } \mathbf{B}. \quad (4.70)$$

Evaluating this expression for the data on India and Japan, we see that the sensitivity of  $\boldsymbol{\eta}^\dagger$  shows a pattern similar to that of the sensitivity of  $V(\boldsymbol{\eta})$  (Fig. 4.2), confirming that these indices are measuring similar aspects of disparity in longevity.

In particular, they show the existence of a critical age, before which reductions in mortality reduce disparity and after which they have the opposite effect. Zhang and Vaupel (2009) showed that this critical age, which they describe as separating “early” from “late” deaths is a general property of  $\boldsymbol{\eta}^\dagger$ . Although the details depend on which index of disparity one uses, the existence of a critical age separating positive and negative sensitivities is also a property of other measures of variation in longevity (Van Raalte and Caswell 2013). Vaupel et al. (2011) have used the critical age to decompose historical changes in lifespan disparity into components due to early and late mortality.

## 4.5 A Time-Series LTRE Decomposition: Life Disparity

The LTRE decomposition analysis in Sect. 2.9 can be used to decompose time series such as these into their components. We apply it here to calculate the contributions, to a long trajectory of changes in  $\eta^\dagger$ , of changes in early and late mortality.

Suppose that some demographic outcome  $\xi(t)$  (dimension  $s \times 1$ ) is measured as a function of a parameter vector  $\theta$  (dimension  $p \times 1$ ), at times  $1, 2, \dots, T$ . The changes in  $\xi(t)$  over time result from the changes in the parameters,

$$\Delta \xi(t) = \xi(t+1) - \xi(t) \quad (4.71)$$

$$\Delta \theta(t) = \theta(t+1) - \theta(t) \quad (4.72)$$

The decomposition analysis for such sequences was introduced as a “regression LTRE” method in the context of ecotoxicology and response to environmental factors (e.g., Caswell 1996; Knight et al. 2009). The same approach was introduced independently by Horiuchi et al. (2008) to decompose differences between two conditions by imagining a continuous path from one to the other.

The analysis starts by considering the change in  $\xi$  over time,

$$\frac{d\xi(t)}{dt} = \frac{d\xi(t)}{d\theta^\top(t)} \frac{d\theta(t)}{dt} \quad (4.73)$$

If the time series is evaluated at discrete times  $t = 1, \dots, T$ , then to first order

$$\Delta \xi(t) \approx \frac{d\xi(t)}{d\theta^\top(t)} \Delta \theta(t) \quad s \times 1 \quad (4.74)$$

The contributions to  $\Delta \xi(t)$  are displayed separately in a contribution matrix

$$\mathbf{C}(t) = \frac{d\xi(t)}{d\theta^\top(t)} \mathcal{D}[\Delta \theta(t)] \quad s \times p \quad (4.75)$$

the  $(i, j)$  entry of  $\mathbf{C}(t)$  is the contribution of  $\Delta \theta_j(t)$  to  $\Delta \xi_i(t)$ . The contributions additive over time, so the contributions of all the changes, integrated from  $t_1$  to  $t_2$ , are given by the entries of

$$\mathbf{C}(t_1, t_2) = \sum_{t=t_1}^{t_2} \mathbf{C}(t) \quad (4.76)$$

Suppose the dependent variable is  $\xi = \eta^\dagger$  and the parameter vector is  $\theta = \mu$ .

At each time and for each age, we aggregate the contributions from early and late mortality. Let  $\mathbf{X}$  be an indicator matrix whose entries define whether a particular entry of  $\mathbf{C}(t)$  is to be counted as early or late:

$$x_{ij} = \begin{cases} 1 & \theta_j \text{ contributes to } \Delta\xi_i \\ 0 & \text{otherwise} \end{cases} \quad (4.77)$$

Then

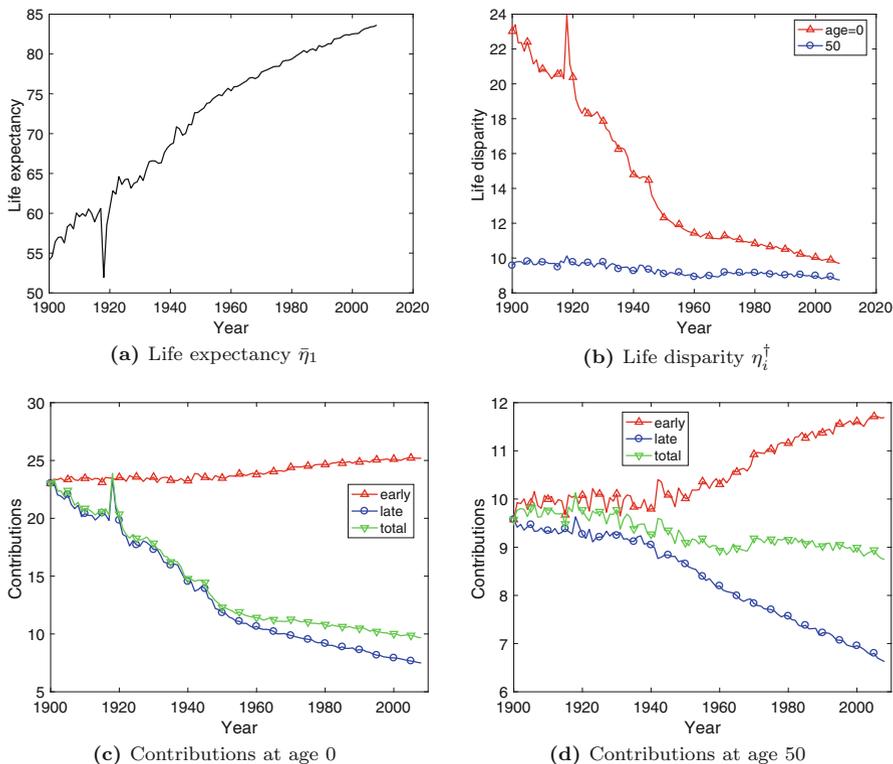
$$\mathbf{c}(t) = (\mathbf{C}(t) \circ \mathbf{X}) \mathbf{1} \quad (4.78)$$

is a vector giving the contributions to the change in  $\xi$  from the parameters chosen in  $\mathbf{X}$ . Defining  $\mathbf{X}_{\text{early}}$  and  $\mathbf{X}_{\text{late}}$  gives changes at time  $t$  due to early and late mortality. The LTRE analysis is then

$$\mathbf{c}_{\text{early}}(t_1, t_2) = \sum_{t_1}^{t_2} \mathbf{c}_{\text{early}}(t) \quad (4.79)$$

and similarly for  $\mathbf{c}_{\text{late}}(t_1, t_2)$ .

As an example, Fig. 4.3a, b shows a time series of life expectancy (increasing from about 40–80 years between 1800 and 2010) and life disparity for Swedish



**Fig. 4.3** (a) Historical trends in life expectancy at birth from 1800 to 2010. (b) Historical trends in life disparity (mean years of life lost due to mortality) for ages 0 and 50 years. (c) Contributions from early and late mortality improvement to the change in disparity at age 0. (d) The contributions for disparity at age 50. (Data for Swedish females, from the Human Mortality Database)

females, based on data from Human Mortality Database (2016). As in most developed countries, life disparity at birth dropped dramatically from 1850 to about 1950 (e.g., Edwards 2011; Vaupel et al. 2011). Declines at later ages were less dramatic, and remaining life disparity conditional on survival to age 50 has been almost flat (Engelman et al. 2014). How did changes in early and late mortality contribute to these patterns?

Figure 4.3c, d show the cumulative sums of the contributions  $c_{\text{early}}$  and  $c_{\text{late}}$ , and their total, for ages 0 and 50. The decline in life disparity at birth was driven almost completely by improvements in early mortality, which completely overshadowed a small increase in disparity that was generated by improvements in late life mortality. The picture for remaining life disparity at age 50 is different: the contributions from changes in early and late life mortality almost completely cancel each other out. These patterns, looking at the details of a single time series, agree with the much more general exploration of multiple countries, using a different approach, by Vaupel et al. (2011).

The accuracy of the decomposition can be evaluated by comparing the time series calculated from the total contributions, as shown in Fig. 4.3c, d, with the observed series, as shown in Fig. 4.3b. The agreement is extremely close; the LTRE decomposition captures the end result of the historical changes from 1800 to 2010 with an error of less than 0.1%.

## 4.6 Conclusion

This chapter and Chap. 3 contain examples of different approaches to the sensitivity analysis, of population growth rate and longevity, respectively. The power and flexibility of matrix calculus methods is apparent: the models are not restricted to age- or stage-classification, the absorbing states may be a single category of death or some more diverse set, the demographic outcomes are not limited to expectations, and the independent variables, the parameters that are being perturbed, can be anything of interest. The only requirement is that a chain of functional dependence can be followed: the outcome  $\xi$  depends on  $\mathbf{U}$ , which depends on  $\mathbf{p}$ , which depends on  $\boldsymbol{\mu}$ , . . . and so on. Mortality might depend on health status, which might depend on income level, which might depend on education, . . . , and so on. The sensitivity of  $\xi$  to any of these parameters is a application of the chain rule.

## Bibliography

- Canudas Romo, V. 2003. Decomposition methods in demography. Population Studies, Rozenberg Publishers, Amsterdam, Netherlands.
- Caswell, H. 1996. Analysis of life table response experiments. II. Alternative parameterizations for size- and stage-structured models. Ecological Modelling 88:73–82.

- Caswell, H. 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*. 2nd edition. Sinauer Associates, Sunderland, MA.
- Caswell, H., 2006. Applications of Markov chains in demography. Pages 319–334 in *MAM2006: Markov Anniversary Meeting*. Bosen Books, Raleigh, North Carolina.
- Caswell, H. 2009. Stage, age and individual stochasticity in demography. *Oikos* **118**:1763–1782.
- Caswell, H. 2011. Beyond  $R_0$ : Demographic models for variability of lifetime reproductive output. *PLoS ONE* **6**:e20809.
- Caswell, H., and F. A. Kluge. 2015. Demography and the statistics of lifetime economic transfers under individual stochasticity. *Demographic Research* **32**:563–588.
- Caswell, H., and V. Zarulli. 2018. Matrix methods in health demography: a new approach to the stochastic analysis of healthy longevity and DALYs. *Population Health Metrics* **16**:8.
- Devleesschauwer, B., A. H. Havelaar, C. M. De Noordhout, J. A. Haagsma, N. Praet, P. Dorny, L. Duchateau, P. R. Torgerson, H. Van Oyen, and N. Speybroeck. 2014. DALY calculation in practice: a stepwise approach. *International Journal of Public Health* **59**:571–574.
- Edwards, R. D. 2011. Changes in world inequality in length of life: 1970–2000. *Population and Development Review* **37**:499–528.
- Edwards, R. D., and S. Tuljapurkar. 2005. Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review* **31**:645–674.
- Engelman, M., H. Caswell, and E. M. Agree. 2014. Why do lifespan variability trends for the young and old diverge? *Demographic Research* **30**:1367–1396.
- Feichtinger, G. 1971. *Stochastische Modelle Demographischer Prozesse*. Lecture notes in operations research and mathematical systems, Springer-Verlag, Berlin, Germany.
- GBD 2016 DALYs and HALE Collaborators. 2017. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**:1260–1344.
- Hoem, J. M. 1969. Markov chain models in life insurance. *Sonderdruck aus Blatter der Deutschen Gesellschaft fur Versicherungsmathematik* **9**:91–107.
- Horiuchi, S., J. R. Wilmoth, and S. D. Pletcher. 2008. A decomposition method based on a model of continuous change. *Demography* **45**:785–801.
- Horvitz, C. C., and S. Tuljapurkar. 2008. Stage dynamics, period survival, and mortality plateaus. *American Naturalist* **172**:203–215.
- Howard, R. A. 1960. *Dynamic programming and Markov processes*. Wiley, New York, New York.
- Human Mortality Database. 2016. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). [www.mortality.org](http://www.mortality.org) URL [www.mortality.org](http://www.mortality.org).
- Iosifescu, M. 1980. *Finite Markov Processes and Their Applications*. Wiley, New York, New York.
- Irwin, J. 1949. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *The Journal of Hygiene* **47**:188–189.
- Kemeny, J. G., and J. L. Snell. 1976. *Finite Markov Chains*. Second edition. Undergraduate Texts in Mathematics, Springer-Verlag, New York, New York, USA.
- Keyfitz, N. 1971. Linkages of intrinsic to age-specific rates. *Journal of the American Statistical Association* **66**:275–281.
- Knight, T., H. Caswell, and S. Kalisz. 2009. Population growth rate of a common understory herb decreases non-linearly across a gradient of deer herbivory. *Forest Ecology and Management* **257**:1095–1103.
- Pollard, J. H. 1982. The expectation of life and its relationship to mortality. *Journal of the Institute of Actuaries* **109**:225–240.
- Steinsaltz, D., and S. N. Evans. 2004. Markov mortality models: implications of quasistationarity and varying initial distributions. *Theoretical Population Biology* **65**:319–337.
- Tuljapurkar, S., and C. C. Horvitz. 2006. From stage to age in variable environments: life expectancy and survivorship. *Ecology* **87**:1497–1509.

- van Daalen, S., and H. Caswell. 2015. Lifetime reproduction and the second demographic transition: Stochasticity and individual variation. *Demographic Research* **33**:561–588.
- van Daalen, S. F., and H. Caswell. 2017. Lifetime reproductive output: individual stochasticity, variance, and sensitivity analysis. *Theoretical Ecology* **10**:355–374.
- Van Raalte, A. A., and H. Caswell. 2013. Perturbation analysis of indices of lifespan variability. *Demography* **50**:1615–1640.
- Vaupel, J. W. 1986. How change in age-specific mortality affects life expectancy. *Population Studies* **40**:147–157.
- Vaupel, J. W., and V. Canudas Romo. 2003. Decomposing change in life expectancy: a bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. *Demography* **40**:201–216.
- Vaupel, J. W., Z. Zhang, and A. A. van Raalte. 2011. Life expectancy and disparity: an international comparison of life table data. *BMJ Open* **1**:e000128.
- Wilson, E. B. 1938. The standard deviation of sampling for life expectancy. *Journal of the American Statistical Association* **33**:705–708.
- Zhang, Z., and J. W. Vaupel. 2009. The age separating early deaths from late deaths. *Demographic Research* **20**:721–730.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

