



*Demographic Research* a free, expedited, online journal  
of peer-reviewed research and commentary  
in the population sciences published by the  
Max Planck Institute for Demographic Research  
Konrad-Zuse Str. 1, D-18057 Rostock · GERMANY  
[www.demographic-research.org](http://www.demographic-research.org)

---

## ***DEMOGRAPHIC RESEARCH***

**VOLUME 23, ARTICLE 26, PAGES 737-748  
PUBLISHED 12 OCTOBER 2010**

<http://www.demographic-research.org/Volumes/Vol23/26/>  
DOI: 10.4054/DemRes.2010.23.26

*Formal Relationships 11*

### **Attrition in heterogeneous cohorts**

**James W. Vaupel**

**Zhen Zhang**

This article is part of the Special Collection "Formal Relationships".  
Guest Editors are Joshua R. Goldstein and James W. Vaupel.

© 2010 *James W. Vaupel & Zhen Zhang.*

*This open-access work is published under the terms of the Creative Commons Attribution NonCommercial License 2.0 Germany, which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/by-nc/2.0/de/>*

## Table of Contents

|     |   |     |
|-----|---|-----|
| 1   | Relationship  | 738 |
| 2   | Proof   | 739 |
| 3   | Discussion  | 740 |
| 4   | Applications  | 740 |
| 4.1 | The shape of the hazard of attrition in heterogeneous populations | 740 |
| 4.2 | Estimation of the rate of aging in a heterogeneous cohort         | 742 |
|     | References  | 746 |

## **Attrition in heterogeneous cohorts**

**James W. Vaupel**<sup>1, 3</sup>

**Zhen Zhang**<sup>2, 3</sup>

### **Abstract**

In a heterogeneous cohort, the change with age in the force of mortality or some other kind of hazard or intensity of attrition depends on how the hazard changes with age for the individuals in the cohort and on how the composition of the cohort changes due to the loss of those most vulnerable to attrition. Here we prove that the change with age for the cohort equals the average of the change in the hazard for the individuals in the cohort minus the variance in the hazard across individuals. The variance captures the compositional change. This very general and remarkably elegant relationship can be applied to understand and to analyze changes with age in many kinds of demographic hazards, including, e.g., the life-table aging rate or the intensity of first births.

---

<sup>1</sup> Max Planck Institute for Demographic Research. E-mail: [jwv@demogr.mpg.de](mailto:jwv@demogr.mpg.de).

<sup>2</sup> Max Planck Institute for Demographic Research. E-mail: [zhang@demogr.mpg.de](mailto:zhang@demogr.mpg.de).

<sup>3</sup> We thank Maxim Finkelstein and Michel Guillot for constructive comments.

## 1. Relationship

Let  $\mu(x, z)$  denote the hazard of attrition (e.g., the force of mortality) at age  $x$  for a cohort of individuals in a subpopulation with characteristic  $z$ ; the subpopulation could consist of a single individual. Let  $g(z)$  denote the probability density function of  $z$  at initial age zero, let  $s(x, z) = \exp[-\int_0^x \mu(a, z)da]$  denote the survival function for the subpopulation with characteristic  $z$  and let  $\bar{s}(x) = \int_{-\infty}^{\infty} s(x, z)g(z)dz$  denote the survival function for the population as a whole. Then

$$(1a) \quad \dot{\bar{\mu}}(x) = \bar{\dot{\mu}}(x) - \sigma_{\mu}^2(x),$$

where  $\bar{\mu}(x) = \int_{-\infty}^{\infty} \mu(x, z)s(x, z)g(z)dz/\bar{s}(x)$  is the average hazard of attrition for the population at age  $x$  as a whole,  $\dot{\bar{\mu}}(x) = \frac{d\bar{\mu}(x)}{dx}$  is the change in this hazard with age for the population as a whole at age  $x$ ,  $\dot{\mu}(x, z) = \frac{d\mu(x, z)}{dx}$  is the change in the hazard of attrition with age for subpopulation with characteristic  $z$  at age  $x$ ,  $\bar{\dot{\mu}}(x) = \int_{-\infty}^{\infty} \dot{\mu}(x, z)s(x, z)g(z)dz/\bar{s}(x)$  is the average change in the hazard of attrition for subpopulations at age  $x$ , and  $\sigma_{\mu}^2(x) = \int_{-\infty}^{\infty} \mu^2(x, z)s(x, z)g(z)dz/\bar{s}(x) - \bar{\mu}^2(x)$  is the variance at age  $x$  in the hazard of attrition across subpopulations.

A relationship similar to (1a) holds for the case of a characteristic that is discrete. Let  $\pi_z(x)$  be the proportion of the subpopulation with characteristic  $z$  surviving at age  $x$ , with  $\pi_z(x) = \pi_z(0) \exp[-\int_0^x \mu_z(t)dt]$ , where  $\mu_z(x)$  is the hazard of attrition of the subpopulation with characteristic  $z$  at age  $x$  and  $\pi_z(0)$  is the proportion of the subpopulation with characteristic  $z$  at the initial age. Then the average of mortality of the whole population is given by

$$\bar{\mu}(x) = \frac{\sum_z \mu_z(x)\pi_z(x)}{\sum_z \pi_z(x)}$$

and

$$(1b) \quad \dot{\bar{\mu}}(x) = \bar{\dot{\mu}}(x) - \sigma_{\mu}^2(x),$$

where  $\bar{\dot{\mu}}(x) = \frac{\sum_z \dot{\mu}_z(x)\pi_z(x)}{\sum_z \pi_z(x)}$  and  $\sigma_{\mu}^2(x) = \frac{\sum_z \mu_z^2(x)\pi_z(x)}{\sum_z \pi_z(x)} - \bar{\mu}^2(x)$ .

Equation (1b) is essentially the same as (1a) and can be used to study attrition in a population that is heterogeneous with regard to a discrete characteristic.

In short, the change in the average equals the average change minus the variance.

## 2. Proof

Because  $\bar{\mu}(x) = \int_{-\infty}^{\infty} \mu(x, z)s(x, z)g(z)dz/\bar{s}(x)$ , it follows that

$$\dot{\bar{\mu}}(x) = \frac{\int_{-\infty}^{\infty} \{\dot{\mu}(x, z)s(x, z) + \mu(x, z)\dot{s}(x, z)\}g(z)dz}{\bar{s}(x)} - \bar{\mu}(x)\frac{\dot{\bar{s}}(x)}{\bar{s}(x)}.$$

Substituting  $\dot{s}(x, z) = -\mu(x, z)s(x, z)$  and  $\dot{\bar{s}}(x) = -\bar{\mu}(x)\bar{s}(x)$  and simplifying yields

$$\begin{aligned}\dot{\bar{\mu}}(x) &= \bar{\mu}(x) - \frac{\int_{-\infty}^{\infty} \mu^2(x, z)s(x, z)g(z)dz}{\bar{s}(x)} + \bar{\mu}^2(x) \\ &= \bar{\mu}(x) - \sigma_{\mu}^2(x).\end{aligned}\quad \text{Q.E.D.}$$

Equation (1b) can be proven in a similar manner. Simple calculus yields

$$\dot{\bar{\mu}}(x) = \frac{\sum_z (\dot{\mu}(x)\pi_z(x) + \mu_z(x)\dot{\pi}_z(x))}{\sum_z \pi_z(x)} - \bar{\mu}(x)\frac{\sum_z \dot{\pi}_z(x)}{\sum_z \pi_z(x)}.$$

Because  $\mu_z(x) = -\dot{\pi}_z(x)/\pi_z(x)$ , it follows that  $\dot{\pi}_z(x) = -\mu_z(x)\pi_z(x)$ . Substituting this expression into the above equation leads to

$$\dot{\bar{\mu}}(x) = \bar{\mu}(x) - \sigma_{\mu}^2(x)\quad \text{Q.E.D.}$$

Equation (1a) - and with minor modifications equation (1b) - can also be proven as a special case of the more general equation

$$(2) \quad \dot{\bar{v}}(x) = \bar{v}(x) + Cov(v(x), r(x)).$$

In (2),  $v(x, z)$  gives the value of some function of age or time  $x$  in the subpopulation with characteristic  $z$ , the proportion of individuals at age or time  $x$  in the subpopulation with characteristic  $z$  is given by the probability density function  $g(x, z)$ ,

$$\begin{aligned}\bar{v}(x) &= \int_{-\infty}^{\infty} v(x, z)g(x, z)dz, & \dot{\bar{v}}(x) &= \frac{d\bar{v}(x)}{dx}, & \dot{v}(x, z) &= \frac{dv(x, z)}{dx} & \text{and} \\ \bar{\dot{v}}(x) &= \int_{-\infty}^{\infty} \dot{v}(x, z)g(x, z)dz.\end{aligned}$$

The covariance is given by

$$Cov(v(x), r(x)) = \int_{-\infty}^{\infty} v(x, z)r(x, z)g(x, z)dz - \bar{v}(x)\bar{r}(x),$$

where  $r(x, z) = \frac{dg(x, z)/dx}{g(x, z)}$  is the relative derivative of the density function and  $\bar{r}(x) = \int_{-\infty}^{\infty} r(x, z)g(x, z)dz$ . To prove (1a), substitute  $\mu$  for  $v$  in (2) and substitute  $s(x, z)g(z)/\bar{s}(x)$  for  $g(x, z)$ . It readily can be shown that  $r(x, z) = -\mu(x, z) + \bar{\mu}(x)$ . It follows that  $Cov(v(x), r(x)) = -\sigma_{\mu}^2(x)$ .

### 3. Discussion

Equation (1) is very general. In the frailty model developed by Vaupel, Manton, and Stallard (1979), an individual's frailty is a relative-risk that is fixed for life, and in many applications frailty is assumed to be gamma distributed. Equation (1), however, holds regardless of the distribution of  $z$  and even if  $z$  is not a relative-risk (proportional-hazard) but  $z$  affects the hazard of attrition in some complicated manner. The value of  $z$  can be positive, zero or negative. Every individual in the population could have a unique value of  $z$  that determines the mortality trajectory over age for that individual. Depending on the individual, such trajectories could rise or fall or rise and then fall, etc. Changing-frailty models in which an individual's risk varies with age, perhaps stochastically, can be captured by defining  $z$  as an index of every possible trajectory.

Both (1) and (2) were presented at an annual meeting of the Population Association of America (Vaupel 1992). A proof of (2) is given by Vaupel and Canudas-Romo (2002). A version of (2) without the first term on the right was derived by Preston, Himes and Eggers (1989). A discrete version of (2) was devised by Price (1970) to analyze genetic change. Cohen (1971) independently derived the same result. Coulson and Tuljapurkar (2008) extend the so-called Price equation.

## 4. Applications

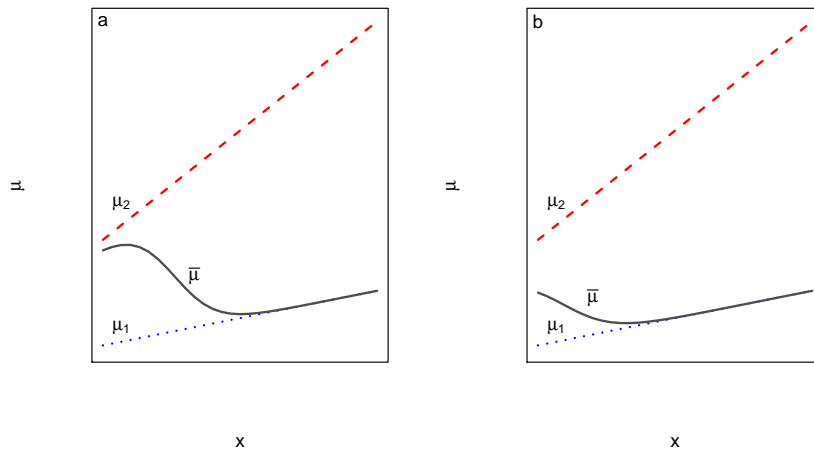
### 4.1 The shape of the hazard of attrition in heterogeneous populations

Consider the simplest case where the hazard of attrition for individuals is constant over age but varies across individuals. Then  $\dot{\mu}(x) = 0$  and  $\bar{\dot{\mu}}(x) = 0$ , resulting in  $\dot{\bar{\mu}}(x) = -\sigma_{\mu}^2(x) < 0$  for all  $x$ . That is, the hazard of attrition for the overall population decreases with age.

If the hazard of attrition for subpopulations is decreasing with age (i.e.  $\dot{\mu}(x, z) < 0$  and  $\bar{\dot{\mu}}(x) < 0$ ), the hazard of attrition for the population as a whole must decrease with age, because both terms on the right hand side of (1),  $\bar{\dot{\mu}}(x)$  and  $-\sigma_{\mu}^2(x)$ , are negative. This result is well known in reliability theory (e.g., Barlow 1985, Proschan 1963).

The case of increasing mortality for individuals, however, is not so simple: the average hazard of attrition for the population as a whole could either increase or decrease at age  $x$  depending on the value of  $\bar{\dot{\mu}}(x)$  vs.  $\sigma_{\mu}^2(x)$ . As an example, consider a population composed of two subpopulations with forces of attrition  $\mu_1(x) = c_1t + a_1$  and  $\mu_2(x) = c_2t + a_2$ , with  $0 < c_1 < c_2$  and  $0 < a_1 < a_2$ . Then the hazard of attrition for the overall population can decrease over some ages and increase over other ages, as shown in Figures 1a and 1b.

**Figure 1: The divergence of the forces of attrition for the population as a whole and the two subpopulations.**



Note: The blue dotted line represents  $\mu_1(x) = c_1t + a_1$  with  $c_1 = .5$  and  $a_1 = .1$ , the red dashed line  $\mu_2(x) = c_2t + a_2$  with  $c_2 = 2$  and  $a_2 = 3$ . The solid line represents the force of attrition for the population as a whole. The only difference between the two figures is that the share of the subpopulation with  $\mu_1(x)$  at initial age zero is .1 in Figure 1a and .5 in Figure 1b. The graph in Figure 1a is similar to a graph in Vaupel and Yashin (1985) and the graph in Figure 1b is similar to graphs in Finkelstein (2009) and Block, Li, and Savtis (2003).

The pattern of attrition for the population as a whole can differ from the patterns for subpopulations or individuals in a variety of ways, as illustrated by Vaupel and Yashin (1985). The above examples illustrate how equation (1) can shed light on patterns of attrition in heterogeneous populations.

## 4.2 Estimation of the rate of aging in a heterogeneous cohort

A more specific application of equation (1) is in studies of the rate of aging, measured as the relative change with age in the force of mortality. Note that dividing both sides of (1) by  $\bar{\mu}(x)$  yields the relative change in the force of mortality for the whole population:

$$(3) \quad \bar{b}(x) = \frac{d\bar{\mu}(x)}{\bar{\mu}(x)dx} = \frac{\bar{\dot{\mu}}(x)}{\bar{\mu}(x)} - \frac{\sigma_{\bar{\mu}}^2(x)}{\bar{\mu}(x)} = b^\dagger(x) - \bar{\mu}(x)CV_{\bar{\mu}}(x)$$

where

$$b^\dagger(x) = \frac{\bar{\dot{\mu}}(x)}{\bar{\mu}(x)} = \frac{\int_{-\infty}^{\infty} \frac{d\mu(x,z)/dx}{\mu(x,z)} \mu(x,z)s(x,z)g(z)dz / \bar{s}(x)}{\int_{-\infty}^{\infty} \mu(x,z)s(x,z)g(z)dz / \bar{s}(x)}$$

is the deaths-weighted rate of aging (because the density of deaths is given by the product of  $\mu$  and  $s$ ) and  $CV_{\bar{\mu}}(x) = \sigma_{\bar{\mu}}^2(x)/\bar{\mu}^2(x)$  is the coefficient of variation at age  $x$ .

Suppose the force of mortality is given by  $\mu(x,z) = za \exp(bx) + c$  with frailty  $z$  at initial age zero described by a gamma distribution with mean 1 and variance  $\gamma$ . Then, results in Vaupel, Manton and Stallard (1979) imply that:

$$(4) \quad \bar{\mu}(x) = \bar{z}(x)(\mu(x,1) - c) + c = \frac{a \exp(bx)}{1 + \frac{a\gamma}{b}(\exp(bx) - 1)} + c,$$

where  $\bar{z}(x)$  is the mean value of frailty among survivors at age  $x$ . It appears that such a logistic curve may capture the pattern of human mortality from age 35 or so to the most advanced ages (Thatcher, Kannisto and Vaupel 1998, Vaupel 2010). It can be readily shown that  $\bar{\dot{\mu}}(x) = b(\bar{\mu}(x) - c)$ . Furthermore, because (4) can be rearranged as  $\bar{\mu}(x) - c = (\mu(x,1) - c)\bar{z}(x)$ , it follows that  $\sigma_{\bar{\mu}}^2(x) = \gamma(\bar{\mu}(x) - c)^2$ , where  $\gamma = \sigma_z^2(x)/\bar{z}^2(x)$  is the coefficient of variation of the gamma distribution of frailty. Note that an important property of the gamma distribution is that  $\gamma$  is constant, and that when mean frailty at the initial age zero is one, then  $\gamma$  is the variance in frailty at age zero.



These results, substituted in (3), imply:

$$(5) \quad \bar{b}(x) = \frac{\dot{\bar{\mu}}(x)}{\bar{\mu}(x)} = b \left( 1 - \frac{c}{\bar{\mu}(x)} \right) - \gamma \left( 1 - \frac{c}{\bar{\mu}(x)} \right) (\bar{\mu}(x) - c),$$

where the relative derivative  $\bar{b}(x)$  is the lifetable aging rate (Horiuchi and Coale 1990, Horiuchi and Wilmoth 1997, 1998).

The curve in (4) asymptotically approaches a limit of  $\bar{\mu}_* = b/\gamma + c$ . In recent decades for countries with high life expectancies, it appears that this limit is closely approached by age 110, such that the force of mortality after 110 is essentially constant at a value of about .7, which implies an annual probability of death of about 50% (Gampe 2010). In any case, at the limit (4) indicates:

$$(6) \quad \bar{b}(x) = 0 = b \left( 1 - \frac{c}{\bar{\mu}_*} \right) - \gamma \left( 1 - \frac{c}{\bar{\mu}_*} \right) (\bar{\mu}_* - c),$$

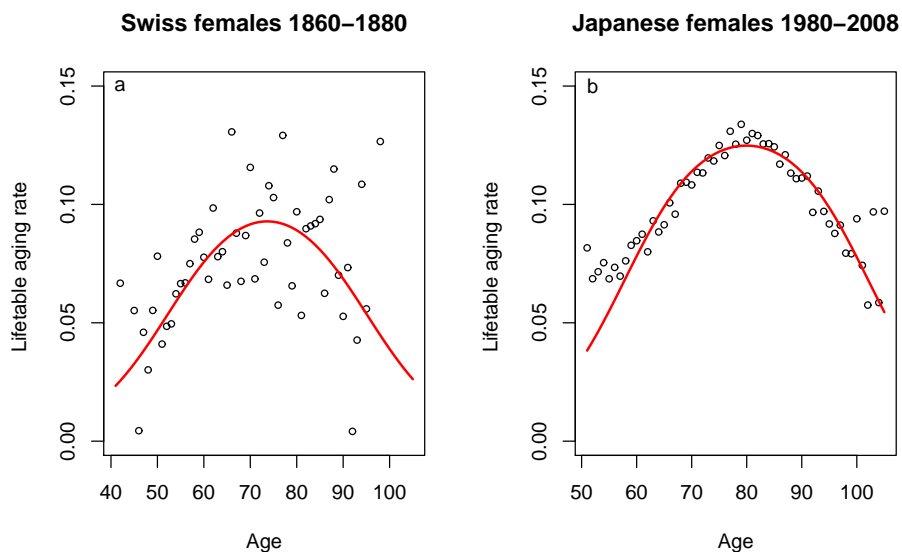
which in turn implies  $\gamma = b/(\bar{\mu}_* - c)$ . Substituting this result in (5) yields:

$$(7) \quad \bar{b}(x) = b \left( 1 - \frac{c}{\bar{\mu}(x)} \right) \left( \frac{\bar{\mu}_* - \bar{\mu}(x)}{\bar{\mu}_* - c} \right).$$

Equation (7) shows how and why the lifetable aging rate  $\bar{b}(x)$  differs from the rate of increase in mortality  $b$ , which is constant for all the members of the population. The first adjustment to  $b$  reduces the value of  $\bar{b}(x)$  at younger ages when  $c$  accounts for a substantial share of total mortality. The second adjustment to  $b$  reduces the value of  $\bar{b}(x)$  at older ages as the force of mortality approaches its maximum value  $\bar{\mu}_*$ . If  $c$  is small compared with this maximum value, then the adjustment is approximately equal to  $(1 - \frac{\bar{\mu}(x)}{\bar{\mu}_*})$ . At the maximum value  $\bar{b}(x)$  equals zero. Note that because  $\bar{\mu}(x) \geq c$ , both factors adjusting  $b$  are less than or equal to one. Hence,  $\bar{b}(x) \leq b$ , and sometimes  $\bar{b}(x)$  can be much lower than  $b$ .

For Japanese females in 1980-2008, the force of mortality from age 50 to 105 can be serviceably approximated, using methods of maximum likelihood estimation, by (4) with initial age zero corresponding to chronological age 50, with  $a = .0007$  at the initial age,  $b = .14$ ,  $c = .002$  and  $\gamma = .19$ ; these values imply that  $\bar{\mu}_* = .73$ . For Swiss females in 1860-1880, the parameter estimates are  $a = .0018$ ,  $b = .11$ ,  $c = .007$ ,  $\gamma = .15$  and  $\bar{\mu}_* = .76$ . The trajectory of  $\bar{b}(x)$  estimated by (7) resembles the trend of change in the empirical  $\bar{b}(x)$  calculated directly from the data, as shown in Figures 2a and 2b.

**Figure 2:** Lifetable aging rate for Swiss females in 1860-1880 (a) and for Japanese females in 1980-2008 (b).



*Note:* The black circles represent empirical lifetable aging rates calculated directly from the data and the red line represents the lifetable aging rate estimated by (7).

*Source of data:* Human Mortality Database available at <http://www.demogr.mpg.de>.

In addition to applications to mortality, equation (1) can also be used in studies of other kinds of attrition with age in a heterogeneous cohort. The cohort could pertain to childless women who drop out when they have a child, single women who marry, men who become widowers or who divorce, natives of a country who leave if they emigrate, healthy people who drop out when they get sick with some kind of disease, sick people who drop out when they regain their health, employed people who lose their jobs, unemployed people who take a job, automobiles bought new that are sold to a second owner, etc.

Consider, for instance, the attrition of childless women when they give birth. Suppose that in such a cohort, some of the women are married and the rest are not married. Suppose the intensity (i.e., force or hazard) of a first birth is higher for the married women than for the unmarried women, and suppose that the women are at an age  $x$  at which the intensity is rising for both groups. To be specific, suppose that half of the women are in each group, that for the married  $\mu_m(x) = .5$  and  $\dot{\mu}_m(x) = .02$  and that for the unmarried  $\mu_u(x) = .1$  and  $\dot{\mu}_u(x) = .02$ . Then  $\bar{\mu}(x) = .02$  and  $\sigma_\mu^2(x) = .04$ . Equation (1) indicates that  $\dot{\bar{\mu}}(x) = -.02$ . Hence, even though the intensity of first births is increasing at a pace of 2% for both married and unmarried women, the intensity of first births is falling at this pace for the cohort as a whole.

## References

- Barlow, R.E. (1985). A Bayes explanation of an apparent failure rate paradox. *IEEE Transactions on Reliability* 34(2): 107–108. doi:10.1109/TR.1985.5221964.
- Block, H.W., Li, Y., and Savtis, T.H. (2003). Initial and final behavior of failure rate functions for mixtures and functions. *Journal of Applied Probability* 40(3): 721–740. doi:10.1239/jap/1059060898.
- Cohen, J.E. (1971). Legal abortions, socioeconomic status, and measured intelligence in the United States. *Biodemography and Social Biology* 18(1): 55–63. doi:10.1080/19485565.1971.9987900.
- Coulson, T. and Tuljapurkar, S. (2008). The dynamics of a quantitative trait in an age-structured population living in a variable environment. *The American Naturalist* 172(5): 599–612. doi:10.1086/591693.
- Finkelstein, M.S. (2009). Understanding the shape of the mixture failure rate (with engineering and demographic applications). *Applied Stochastic Models in Business and Industry* 25(6): 643–663. doi:10.1002/asmb.815.
- Gampe, J. (2010). Human mortality beyond age 110. In: H. Maier, J.G.B. Jeune, J.M. Robine, and J.W. Vaupel (eds.) *Supercentenarians*. Heidelberg: Springer: 219–230. doi:10.1007/978-3-642-11520-2\_13.
- Horiuchi, S. and Coale, A.J. (1990). Age patterns of mortality for older women: An analysis using age-specific rate of mortality change with age. *Mathematical Population Studies* 2(4): 245–267. doi:10.1080/08898489009525312.
- Horiuchi, S. and Wilmoth, J.R. (1997). Age patterns of the life-table aging rate for major causes of death in Japan, 1951–1990. *Journal of Gerontology: Biological Sciences* 52A(1): B67–B77. doi:10.1093/gerona/52A.1.B67.
- Horiuchi, S. and Wilmoth, J.R. (1998). Deceleration in the age pattern of mortality at older ages. *Demography* 35(4): 391–412. doi:10.2307/3004009.
- Preston, S.H., Himes, C.L., and Eggers, M. (1989). Demographic conditions responsible for population aging. *Demography* 26(4): 691–704. doi:10.2307/2061266.
- Price, G.R. (1970). Selection and covariance. *Nature* 227: 520–521. doi:10.1038/227520a0.
- Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics* 5(3): 375–383. doi:10.2307/1266340.
- Thatcher, A.R., Kannisto, V., and Vaupel, J.W. (1998). The force of mortality at ages 80

- to 120. In: *Monographs on Population Aging 5*. Odense, Denmark: Odense University Press. [Http://www.demogr.mpg.de](http://www.demogr.mpg.de).
- Vaupel, J.W. (1992). *Analysis of population changes and differences: Methods for demographers, statisticians, biologists, epidemiologists, and reliability engineers*. Paper presented at the PAA Annual Meeting, Denver, Colorado, April 30 - May 2.
- Vaupel, J.W. (2010). Biodemography of human ageing. *Nature* 464: 536–542. [doi:10.1038/nature08984](https://doi.org/10.1038/nature08984).
- Vaupel, J.W. and Canudas-Romo, V. (2002). Decomposing demographic change into direct vs. compositional components. *Demographic Research* 7(1): 1–14. [doi:10.4054/DemRes.2002.7.1](https://doi.org/10.4054/DemRes.2002.7.1).
- Vaupel, J.W., Manton, K.G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3): 439–454. [doi:10.2307/2061224](https://doi.org/10.2307/2061224).
- Vaupel, J.W. and Yashin, A.I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *The American Statistician* 39(3): 176–185. [doi:10.2307/2683925](https://doi.org/10.2307/2683925).

