The impact of population heterogeneity on the age trajectory of neonatal mortality

Jonas Schöley*

Abstract

Background The risk of death declines rapidly over the first month of life. It has been theorized that the fast pace of the decline is explained by hidden population heterogeneity resulting in a mortality selection process whereby the frailest infants leave the population at the fastest rate. A competing explanation situates the rapid mortality decline on the individual level, pointing towards the risky transition of birth and the subsequent adaptation of the newborn to the unfamiliar surroundings.

Objective To estimate heterogeneity in the level and shape of age-specific mortality within a cohort of newborns and to quantify the degree to which mortality selection explains the shape of the average neonatal mortality trajectory.

Methods Given individual-level data on 20,322,147 births and 82,562 neonatal deaths from the 2008-2012 U.S. birth cohort, I calculate life-tables for 252 mutually exclusive strata each defined by a unique combination of observed birth characteristics. Using this information, I characterize the distribution of mortality risk and its evolution over the first 28 days of life and decompose changes in key characteristics of this distribution – the mean, the variance, and the mean-mode ratio – into a mortality selection and a direct component.

Results The average age trajectory of neonatal mortality is highly influenced by a small group of frail newborns and does not reflect the rather flat age effect estimated for the healthy majority of the birth cohort. While the risk decline over the first day of life is substantially influenced by mortality selection, the overall age trajectory is better explained by the convergence of high-risk towards low-risk population strata.

---

*Interdisciplinary Centre on Population Dynamics, University of Southern Denmark. Correspondence: jschoeley@health.sdu.dk. During the writing of this article the author was a guest at the Max-Planck Institute for Demographic Research and funded by a grant from AXA Insurance.

1. Introduction

Never again throughout the existence of a cohort will the hazard of death change as rapidly as it does during the first few weeks following birth. For the 2009–2012 U.S. birth cohort, the risk of death falls tenfold over the first day of life, and again by a factor of ten over the next four weeks. But how reflective is this population-level phenomenon of a single newborn's risk trajectory? Is the period of exceptional risk right after birth part of every human's experience, or have we been led astray by one more of "Heterogeneity's Ruses" (Vaupel and Yashin 1985)? In this paper, I seek to quantify the impact of population heterogeneity on the age trajectory of neonatal mortality based on *observed* mortality differences in a cohort of U.S. born infants.

Populations in which members are heterogeneous concerning their risk of death are subjected to mortality selection, which may be defined as the changing composition of a cohort over age due to heterogeneous mortality. As mortality is nothing but the rate of leaving a population due to death, with time, the proportion of low-mortality strata will increase, whereas strata with high mortality will become less prevalent. In consequence, any observed change of a cohorts trait over age (such as income, health, risk of death) may either be the result of within-stratum trait changes or result from a change in the cohort's composition along those strata over age due to mortality selection.

A rich (bio)demographic literature exists for the special case of mortality selection driving the shape of the population level hazard of death over age thus explaining phenomena such as late-life mortality plateaus (e.g., Beard 1959; Vaupel et al. 1979; Vaupel and Carey 1993; Steinsaltz and Wachter 2006; Missov and Vaupel 2015; Colchero and Kiyakoglu 2019), the adolescent "accident hump" (Remund 2015), declining mortality following surgery (Hougaard 1986) or the age pattern of early life mortality (Vaupel and Yashin 1983; Hsieh 1985; Trussell and Richards 1985; Avraam et al. 2014). All of the aforementioned literature features frailty models as a means to formalize, understand, and estimate the impact of population heterogeneity on the age trajectory of mortality. In these models, an individual's risk of death depends on a random quantity coined *frailty* (Vaupel et al. 1979) that, on account of being unobserved, may be thought of as hidden heterogeneity between members of a cohort. Frailty is most commonly expressed in a proportional hazards framework where the hazard of death at age x of an individual with frailty z is given by $h(x|z) = zh_0(x)$ and $h_0(x)$ is a *baseline hazard* shared by all members of the population. An expression for the population/marginal/unconditional hazard $\bar{h}(x)$ can be derived[1] by assuming a distribution of survival times for the case where $z = 1$ with corresponding hazard h_0 and a distribution for the frailties at age $x = 0$. It is then possible to fit $\bar{h}(x)$ to observed survival times or a life table and – given the parameters of the fit and an array of formal relationships (Vaupel and Yashin 1983; Vaupel and Missov 2014) – to determine how hidden heterogeneity acts in the population under investigation. Herein lies one central attraction of frailty models: they allow inference about population heterogeneity even if none has been observed. The quality of the inference then, of course, crucially depends on the adequacy of the assumptions going into the model.

---

[1] An excellent introduction to this technique can be found in Wienke (2011).

Figure 1: Predicted hazard (black) vs. life table mortality rates (gray) for the 2008–2012 U.S. birth cohort. Daily mortality rates over the first 365 days of life are well described by a Gompertz baseline distribution with Hougaard distributed frailties. Via the process of mortality selection, a mixture of log-linear individual level hazards, here drawn for various levels of frailty z, gives rise to a population level hazard with extreme curvature on the log scale. The area of the circles is in rough proportion to the number of deaths each day.
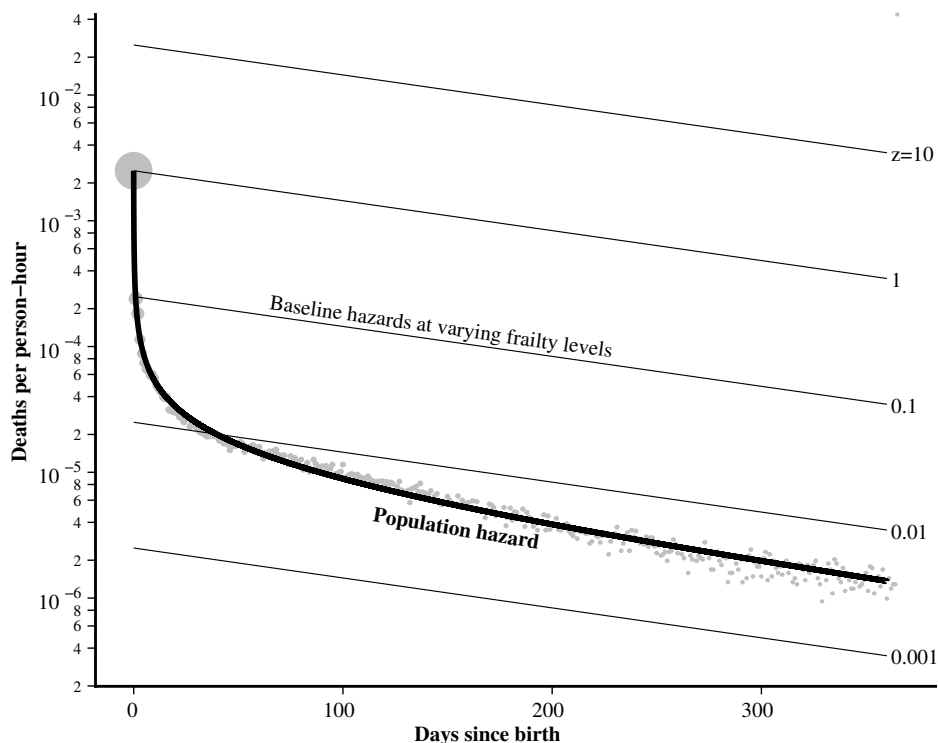
A frailty explanation for the age trajectory of mortality following birth has been put forward multiple times in the literature but never seriously followed up upon. Vaupel and Yashin (1983) modeled the hazard of infant death, assuming a constant baseline hazard and multiplicative Gamma distributed frailty with unity mean and a variance of 500. Such a model implies that the majority of infants at birth are at virtually no risk of death with a small minority of critical cases[2]. Trussell and Richards (1985) demonstrated how sensitive the choice of baseline hazard is when modeling infant mortality via a frailty model, with the Gompertz leading to decreasing and the Weibull to increasing individual-level risk trajectories. Hougaard (1984) hypothesized that the high mortality during the first year of life and the subsequent rapid decline might be the result of an "extreme frailty distribution" upon birth, meaning a distribution with a long right tail, and he proposed the inverse-Gamma distribution as a suitable candidate. Hougaard (1986) introduced a

---

[2] A Gamma distribution with unity mean and variance 500 has more than 98.8 percent of its probability mass below the mean and 0.24 percent at least 100 times above.

particularly flexible family of frailty distributions, including the Gamma, the inverse-Gaussian, and distributions derived from the positive-stable. He fitted a corresponding frailty model to data on time until death after myocardial infarct, a survival scenario not unlike the neonatal case as there is a stressor at time zero with mortality declining quickly and monotonically thereafter.

Indeed the Hougaard frailty model with a Gompertz baseline hazard (see Appendix A for the model specification) gives an excellent fit not only to neonatal mortality but to the entire day-to-day infant life table of the 2009–2012 U.S. birth cohort, capturing the extremely steep decrease in hazard following birth and the subsequent log-linear decline during the post-neonatal period (Figure 1). While by no means a proof of the frailty hypothesis, this simple model shows that a mortality selection explanation is consistent with the dynamics of daily mortality rates observed on the population level.

More recently, the heterogeneous frailty hypothesis was discussed by Levitis (2011) and Levitis and Martínez (2013) as an alternative to evolutionary explanations for "ontogenescence," i.e., the declining hazard of death following birth observed in many species. Via simulation Levitis demonstrated how age-independent individual-level risk leads to age-dependent population mortality rates. In a further variation on the theme, Avraam et al. (2014) captured the decline in mortality from birth to adolescence by employing a discrete frailty model with subgroups featuring either exponentially increasing or constant hazards.

Without data on heterogeneous risks, the hypothesis of mortality selection can only be tested indirectly. While frailty models do provide heterogeneity estimates from population-level data alone, this can only be taken as suggestive evidence for a selection explanation. The estimated amount of heterogeneity depends crucially on the choice of baseline hazard, the choice of frailty distribution, the assumption of fixed versus changing frailty, and the specific way that frailty modulates the baseline hazards (proportional hazards vs. accelerated failure time). These assumptions can not be validated on population-level data as different model specifications yield the same parametric form for the marginal hazard (Trussell and Richards 1985; Hoem 1990; Yashin et al. 2000). I argue that a more convincing case for the impact of population heterogeneity on the age trajectory of mortality can be made by analyzing the distribution of risk in a cohort stratified by *observed* characteristics. For infants, such data is available on birth certificates.

A birth certificate contains a wealth of information that identifies potentially frail newborns. Routinely collected are birth weight, the length of pregnancy, and the Apgar score, an index of the vitality of the child shortly after delivery. The combination of these characteristics delineates hundreds of highly specific subpopulations within a single birth cohort and, based on the distribution of deaths and exposure times across these strata, I analyze the impact of population heterogeneity on the age trajectory of neonatal mortality using both discrete-time life table decomposition methodology and continuous-time hazard modeling in connection with results from formal demography. This approach of describing mortality selection along observed mortality differentials has little precedence as far as human mortality is concerned, an exception being Remund (2015) who, based

on an analysis of French individual level survival data with covariates, found that selection effects contribute to the "accident hump" observed in the hazard trajectory of young adults.

Following the description of the U.S. 2008–2012 birth cohort, I outline a discrete life table approach and a complementary continuous time hazard method to quantify population heterogeneity and mortality selection and its impact on the age trajectory of neonatal mortality. This is followed by results on the distribution of neonatal mortality risks as it changes over age, the heterogeneity of hazard trajectories, and various decompositions of population-level statistics along the age dimension. A discussion of the results concludes the paper.

2. Data

As population heterogeneity drives mortality selection, one can learn about selection effects by estimating the hazard of death across a diverse collection of population strata. Such estimation is possible given the publicly available "NCHS Cohort Linked Birth – Infant Death Data Files" (National Center for Health Statistics 2016) which contain a complete census of births and infant deaths on the territory of the United States (excluding overseas territories) and feature most fields present on the birth and death certificates. The size and detail of the data allow the calculation of neonatal life tables over hundreds of subpopulations, capturing observed heterogeneity in mortality following birth.

To increase the sample size and thus the reliability of the stratum specific mortality estimates, I pool births and deaths across cohorts 2008 to 2012. This leaves a sample size of 20,322,147 births contributing 567,031,738 person-days of exposure to risk over the first 28 days of life, during which 82,562 neonatal deaths were registered.

Following the practice of statistical offices, I compute mortality rates based on *all* registered births and deaths, including cases with missing data on key birth characteristics and unreasonable covariate combinations (e.g., extremely preterm delivery with average birth weight). While this approach challenges any causal interpretation of the determinants of neonatal death, it allows one to capture, in great detail, the observed heterogeneity giving rise to the usual population-level estimates. For this paper, it is of no concern whether the relationship between observed characteristics at birth and risk of death is causal or not as the phenomenon of interest – mortality selection – is induced by any mortality differential between population strata, no matter the cause.

For stratification, I use the information on birth weight, age of gestation at delivery, and 5 minute Apgar score. These variables are routinely recorded upon birth and highly predictive of neonatal death (Pollack et al. 2000; Casey et al. 2001; Park et al. 2018). As the inclusion of further strata only contributes minimal additional heterogeneity compared to what is already captured by the Apgar-birth weight-prematurity trias but considerably increases the computational demands of the model fitting procedure, I opted against it.

The variables are defined and discretized as follows:

Gestation at delivery alias prematurity The number of weeks from conception to delivery commonly estimated by the time since the first day of the mother's last menstrual cycle plus 14 days. Discretized into five groups *Extremely preterm* (earlier than 28 weeks), *Very preterm* (28 to 32 weeks), *Moderate to late preterm* (32 to 37 weeks), *Term or post-term* (37 weeks or later), and *Missing*.

Birth weight The weight of the newborn child measured in the minutes following birth discretized into the five categories *Extremely low* (<1000g), *Very low* (1000–1500g), *Low* (1500–2500g), *Regular or high* (2500g and above) and *Missing*.

5 minute Apgar score A measure of the infant's physical condition five minutes following birth, based on muscle activity, pulse, response to stimulation, skin color, and respiration of the newborn (Apgar 1953). Vitality increases over integers zero to ten, with missings being treated explicitly as a twelfth category.

Table 1 shows births, death counts, and empirical 28-day survival for the complete 2008 to 2012 birth cohort by level of gestation, birth weight, and Apgar score, respectively.

Table 1: Summary statistics of the U.S. 2008–2012 birth cohort.

	Births	(%)	Deaths	(%)	% 28 day survival
Total					
	20,322,147	(100.0)	82,562	(100.0)	99.5
5 minute Apgar score					
0	9,981	(0.1)	5,154	(6.3)	48.3
1	45,091	(0.3)	28,430	(34.5)	36.9
2	30,569	(0.2)	9,839	(12.0)	67.8
3	31,941	(0.2)	4,987	(6.1)	84.3
4	43,408	(0.3)	3,705	(4.5)	91.4
5	72,970	(0.4)	3,781	(4.6)	94.8
6	135,842	(0.7)	4,456	(5.4)	96.7
7	364,256	(1.8)	5,119	(6.3)	98.5
8	2,017,526	(10.0)	5,577	(6.8)	99.7
9	16,746,692	(82.5)	7,869	(9.6)	99.9
10	715,842	(3.6)	242	(0.3)	99.9
(Missing)	108,029	(0.6)	3,403	(4.2)	96.8
Gestation at delivery					
Extremely preterm	149,760	(0.8)	50,606	(61.3)	66.2
Very preterm	246,770	(1.3)	6,990	(8.5)	97.1
Moderate to late preterm	2,029,693	(10.0)	9,484	(11.5)	99.5
Term or post-term	17,870,545	(88.0)	14,493	(17.6)	99.9
(Missing)	25,379	(0.2)	989	(1.2)	96.1
Birth weight					
Extremely low	146,303	(0.8)	52,350	(63.5)	64.2
Very low	151,709	(0.8)	5,692	(6.9)	96.2
Low	1,355,547	(6.7)	10,573	(12.9)	99.2
Regular or high	18,663,662	(91.9)	13,363	(16.2)	99.9
(Missing)	4,926	(0.1)	584	(0.8)	88.1

3. Methods

From neonatal life tables stratified by birth characteristics, I derive the empirical distribution of mortality rates conditioned on time since birth for a cohort of newborns. Age differences in three key characteristics of this distribution – the mean, the variance, and the mean to mode ratio – are then decomposed into components due to mortality selection and due to "direct" effects. Adopting a continuous-time perspective, I estimate hazard trajectories and survival curves for every subpopulation and a corresponding population hazard. By employing the Vaupel-Zhang equality, I calculate the degree to which the between-stratum variance in hazard rates compounds the slope of the population hazard at selected exact ages.

3.1 Stratified infant life-tables

Consider a cohort of $i = 1, \ldots, N$ newborns stratified into subpopulations $k = 1, \ldots, K$ according to the combined discrete characteristics birth weight, age of gestation at birth and five minute Apgar score. Observed for every newborn are a survival time t_{ik} in days since birth censored at day 28 and a binary death indicator δ_{ik}. I partition the survival time into $j = 1, \ldots, J$ non-overlapping daily age intervals $[x_j, x_j + 1)$, where x_j is the start of the jth age interval, and for each interval calculate stratum specific death counts $D_{jk} = \sum_i \delta_{ijk}$, population alive at the beginning of the age interval

$$N_{jk} = \begin{cases} \text{Births}_k & \text{for } j = 1 \\ \text{Births}_k - \sum_{s=1}^{s=j} D_{j=s,k} & \text{for } j > 1 \end{cases},$$

and person-days of exposure $E_{jk} = N_{j+1,k} + a_{jk} D_{jk}$, with a_{jk} as the average time of death of those dying in age interval j. For the first day of life I calculate a_j using information on the proportion of deaths during the first hour of life while for the other days the usual mid-point assumption $a_{jk} = 0.5$ is employed. I then calculate stratum specific life table death rates $m_{jk} = \frac{D_{jk}}{E_{jk}}$ for each age interval and corresponding population level death rates $\overline{m}_j = \frac{\sum_k D_{jk}}{\sum_k E_{jk}}$ and relative exposures $p_{jk} = \frac{E_{jk}}{\sum_k E_{jk}}$. These tabulated counts, exposures, rates and proportions form the basis of all subsequent analyses.

3.2 Discrete-time life table analysis

Decomposing change in mean mortality over age

Using the shorthand $\Delta f_x = f_{x+1} - f_x$ let $\Delta \overline{m}_j$ denote the change in population-level mortality from age interval j to $j+1$. How much of this difference is explained by a change of the subpopulation hazards and how much due to a change in the population composition induced by mortality selection? A straightforward solution to this decomposition problem can be derived by writing $\overline{m}_j = \sum_k p_{jk} m_{jk}$ and applying the product rule for finite

differences (Boole 1880) to the equivalent products $m_{jk}p_{jk} = p_{jk}m_{jk}$ yielding the two expressions

$$\Delta\overline{m}_j = \sum_k p_{jk}\Delta m_{jk} + \sum_k m_{j+1,k}\Delta p_{jk}$$
$$= \sum_k m_{jk}\Delta p_{jk} + \sum_k p_{j+1,k}\Delta m_{jk},$$

which when averaged give the well known Kitagawa decomposition (Kitagawa 1955),

$$\Delta\overline{m}_j = \underbrace{\sum_k \frac{p_{jk} + p_{j+1,k}}{2}\Delta m_{jk}}_{\text{Direct change } \Delta\overline{m}_j^D} + \underbrace{\sum_k \frac{m_{jk} + m_{j+1,k}}{2}\Delta p_{jk}}_{\text{Compositional change } \Delta\overline{m}_j^C}. \qquad (1)$$

The two terms represent the change in the population mortality rate due to changes in the group-specific mortality rates and due to changing group composition. Of particular interest is the ratio $\frac{\Delta\overline{m}_j^C}{\Delta\overline{m}_j}$, which for $\Delta\overline{m}_j^C < 0$ and $\Delta\overline{m}_j < 0$ is the share of the decline in population mortality explained by a compositional shift of the population. If mortality selection is indeed the main driver of the age decline in mortality over the neonatal period, then the aforementioned ratio has to be greater than 0.5.

Decomposing change in mortality rate variance over age

In addition to compounding the effect of age on the average risk of death, mortality selection can lower the variance of mortality rates across population strata as the survivors concentrate in more resilient subgroups. Alternatively, any change in population variance may result from stratum specific mortality rates converging over age. Because the variance of mortality rates across strata k among the survivors in age group j is the weighted average $v_j(x) = \sum_k p_{jk}s_{jk}$, with $s_{jk} = (m_{jk} - \overline{m}_j)^2$, one can follow the same strategy as before to yield the Kitagawa-style decomposition

$$\Delta v_j = \underbrace{\sum_k \frac{p_{jk} + p_{j+1,k}}{2}\Delta s_{jk}}_{\text{Direct change } \Delta v_j^D} + \underbrace{\sum_k \frac{s_{jk} + s_{j+1,k}}{2}\Delta p_{jk}}_{\text{Compositional change } \Delta v_j^C}, \qquad (2)$$

with v_j^D denoting the change in variance due to convergence or divergence of stratum specific mortality rates over age towards the population mean and v_j^C the age-decline in variance due to changing population composition, i.e., mortality selection.

Decomposing change in the ratio of mean to modal mortality over age

If the distribution of death rates in a cohort of infants exhibits positive skewness, mortality selection can reduce the mean-mode ratio of this distribution by "thinning the tail," i.e., by

reducing the relative proportion of very frail subpopulations. However, any such decline may also result from genuine convergence of stratum specific death rates towards the death rates of the most prevalent stratum, the modal mortality rate, formally $\mathcal{M}_j = m_{j,k=r}$ with r such that $p_{j,k=r} = \max(p_{j1}, \ldots, p_{jK})$.

A Kitagawa-style decomposition of the mean-mode ratio $\frac{\overline{m}_j}{\mathcal{M}_j} = \sum_k p_{jk} \frac{m_{jk}}{\mathcal{M}_j}$, shortened to r_j, gives the two components

$$\Delta r_j = \underbrace{\sum_k \frac{p_{jk} + p_{j+1,k}}{2} \Delta \frac{m_{jk}}{\mathcal{M}_j}}_{\text{Direct change } \Delta r_j^D} + \underbrace{\sum_k \frac{\frac{m_{jk}}{\mathcal{M}_j} + \frac{m_{j+1,k}}{\mathcal{M}_{j+1}}}{2} \Delta p_{jk}}_{\text{Compositional change } \Delta r_j^C}, \tag{3}$$

where Δr_j^D captures the changing ratios of stratum specific mortality rates to modal mortality and Δr_j^C referring to changes in the population composition.

3.3 Continuous-time hazard analysis

In the following, I outline a methodology to decompose the age derivative of the population hazard of death into a "direct" component reflecting actual age-effect within the population strata, and a "compositional component" capturing the confounding effect of population heterogeneity.

Estimating stratum specific hazard and survival

Stratum specific hazards during the neonatal period are well captured by the expression

$$h_k(x) = e^{\beta_{0k} + \beta_{1k} \log(x+1) + \beta_{2k} \log^2(x+1)}, \tag{4}$$

which can be interpreted as a Weibull hazard extended by a log-quadratic term over day of life (Appendix B). The curve has the advantage of being linear the log scale, thus allowing it to be fit as a generalized linear mixed/multilevel model, which greatly facilitates a stable estimation of stratum specific hazards. There are in total 252 strata in the population of newborns with each stratum k marking a unique combination of "Prematurity," "Birth weight," and "Apgar score" levels. Fitting the hazard separately to each stratum will result in erratic estimates as some strata do not contain enough observations to inform the model. A solution to this problem is to fit a multilevel model where instead of estimating $\beta_{0k}, \beta_{1k}, \beta_{2k}$ directly one models the coefficient's distribution at different grouping levels.

Given the observed death counts D_{jk} in age group j and stratum k and associated person-days of exposure to risk E_{jk} I fit the model

$$D_{jk} \sim \text{Pois}\left(\lambda_{jk} E_{jk}\right)$$
$$\lambda_{jk} = e^{\beta_{0k} + \beta_{1k} \log(x_{jk}+1) + \beta_{2k} \log^2(x_{jk}+1)}, \tag{5}$$

where λ_{jk} are mortality rates by age group and stratum. For each stratum, a smooth hazard is recovered by evaluating λ_{jk} over a continuous range of ages x as in equation (4).

The stratum specific coefficients $\beta_{0k}, \beta_{1k}, \beta_{2k}$ are sums of baseline coefficients β, prematurity effects β^{Pm}, prematurity-birth weight interactions $\beta^{\text{Pm}\times\text{Bw}}$, and prematurity-birth weight-Apgar interactions $\beta^{\text{Pm}\times\text{Bw}\times\text{Ap}}$ resulting in the multilevel structure

$$\begin{pmatrix} \beta_{0k} \\ \beta_{1k} \\ \beta_{2k} \end{pmatrix} = \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{\substack{\text{lvl 0} \\ \text{baseline coef.}}} + \underbrace{\begin{pmatrix} \beta^{\text{Pm}}_{0,p[k]} \\ \beta^{\text{Pm}}_{1,p[k]} \\ \beta^{\text{Pm}}_{2,p[k]} \end{pmatrix}}_{\substack{\text{lvl 1} \\ \text{deviations by} \\ \text{prematurity}}} + \underbrace{\begin{pmatrix} \beta^{\text{Pm}\times\text{Bw}}_{0,p[k],b[k]} \\ \beta^{\text{Pm}\times\text{Bw}}_{1,p[k],b[k]} \\ \beta^{\text{Pm}\times\text{Bw}}_{2,p[k],b[k]} \end{pmatrix}}_{\substack{\text{lvl 2} \\ \text{deviations by} \\ \text{birth weight given} \\ \text{prematurity}}} + \underbrace{\begin{pmatrix} \beta^{\text{Pm}\times\text{Bw}\times\text{Ap}}_{0,p[k],b[k],a[k]} \\ \beta^{\text{Pm}\times\text{Bw}\times\text{Ap}}_{1,p[k],b[k],a[k]} \\ \beta^{\text{Pm}\times\text{Bw}\times\text{Ap}}_{2,p[k],b[k],a[k]} \end{pmatrix}}_{\substack{\text{lvl 3} \\ \text{deviations by} \\ \text{Apgar given} \\ \text{prematurity and} \\ \text{birth weight}}},$$

where $p[k]$, $b[k]$ and $a[k]$ denote the prematurity, birth weight and Apgar level associated with stratum k. Except for the baseline β's each set of coefficients is assumed to be drawn from a multivariate-Normal distribution with zero mean and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma^2_{\beta_0} & \sigma_{\beta_0\beta_1} & \sigma_{\beta_0\beta_2} \\ \sigma_{\beta_0\beta_1} & \sigma^2_{\beta_1} & \sigma_{\beta_1\beta_2} \\ \sigma_{\beta_0\beta_2} & \sigma_{\beta_1\beta_2} & \sigma^2_{\beta_2} \end{pmatrix},$$

with separate estimates for levels one to three.

Treating the coefficients as realizations from a multivariate-Normal distribution with zero mean acts as a regularizer on the estimated hazard trajectories. In cases where zero or very few deaths are observed, the stratum specific deviations will be "pulled" towards zero with the effect that the estimated hazard resembles the hazard of the next higher hierarchy level (Gelman and Hill 2007). Take as an example the 303 infants delivered "Very preterm" with "Very low" birth weight and given an Apgar score of 10. Because all newborns survived the neonatal period, there is no information available to learn the shape of the underlying hazard. In the multilevel model, the larger population of infants delivered "Very preterm" with "Very low" birth weight will inform the hazard of the Apgar 10 subgroup. If plenty of data is available, then the hazard estimate will follow that data closely.

I fit model (5) as a Generalized Linear Mixed Model using the `lme4` library in R (Bates et al. 2015) and use the fitted coefficients to compute stratum specific hazard trajectories and

survival curves – the basis for the Vaupel-Zhang decomposition – using the expressions given in Appendix C.

Vaupel-Zhang decomposition

The Vaupel-Zhang equality (Vaupel and Zhang 2010) states that in any cohort stratified by some random characteristic K with initial distribution $\pi_k = P(K = k)$ the age derivative of the average hazard of death amongst the survivors at age x can be written as

$$\dot{\overline{h}}(x) = \overline{\dot{h}}(x) - \sigma_h^2(x),$$

where $\overline{\dot{h}}(x)$ denotes average slope and $\sigma_h^2(x)$ the average variance of hazards for the survivors at x. This result can be interpreted as a decomposition of the slope in the population hazard at x into a direct and a compositional component: The first component is based upon the actual slopes of hazards within each stratum k. Only if the population is completely homogeneous with respect to the levels of their hazards is the slope in the population hazard equal to the average slope in the stratum specific hazards. When, however, the variance of hazards is not zero, i.e., the population differs in their risk of death, mortality selection biases the slope in the population level hazard downwards. In other words, the trajectory of the population hazard over age is explained by the average hazard trajectory across population strata compounded by heterogeneity in the level of hazard.

The quantity of interest is the ratio $v(x) = \frac{-\sigma_h^2(x)}{\dot{\overline{h}}(x)}$ which, given that $\dot{\overline{h}}(x) < 0$, is the share of decline in the population hazard at x explained by population heterogeneity in the level of mortality. From the observed distribution of birth characteristics $\pi_k = \frac{\text{Births}_k}{\sum_k \text{Births}_k}$ and the estimated stratum specific hazards $h_k(x)$ and survival $S_k(x)$ I calculate the population survival curve $S(x) = \sum_k \pi_k S_k(x)$ and the stratum distribution conditioned on survival to a given age $\pi_k(x) = \frac{S_k(x)\pi_k}{S(x)}$ which in turn gives rise to the population hazard curve $\overline{h}(x) = \sum_k \pi_k(x)h_k(x)$, and the variance of hazards over population strata $\sigma_h^2(x) = \sum_k \pi_k(x)(h_k(x) - \overline{h}(x))^2$. See Appendix C for a derivation of the stated equalities.

4. Results

4.1 Extreme skewness and Taylor's law in the distribution of mortality

The distribution of mortality rates over population strata conditioned on neonatal age is characterized by the high proportion of lowest-low mortality infants, an extremely long right tail of high-risk subpopulations, and a power-law relationship between expectation and variance (Taylor's law). With more than 72 percent share on the total population, the group of infants who are born on term or post-term with regular or high birth weight and a five minute Apgar score of nine, are the most prevalent stratum throughout the entire neonatal period. Their mortality (the mode of the distribution of mortality risk)

is among the lowest observed and remains relatively constant throughout the first four weeks of life. The quantile of average mortality well demonstrates the extremely long right tail of the mortality risk distribution: On the day of birth, 98 percent of newborns are part of a stratum with a risk of death lower than the population average. Over the next four weeks, the skewness decreases, but the expected value remains a bad measure of centrality (Figure 2B). On a log-log plot, the relationship between mean and variance of the distribution of mortality risk over the first 28 days of life is almost perfectly linear with a slope of 2.5. The fitted regression model reproduces this linearity (Figure 2A).

Figure 2: Mean-variance relationship over day of age (A) and age-specific distribution function (B) of the mortality/hazard rates in a cohort of newborns across birth weight, gestational age, and Apgar score strata. The points in (A) mark mean and variance of the life table mortality rates, whereas the smooth line is the prediction from the Poisson regression. In plot (B), the gray lines refer to the life table estimates and the black lines to the model predictions.
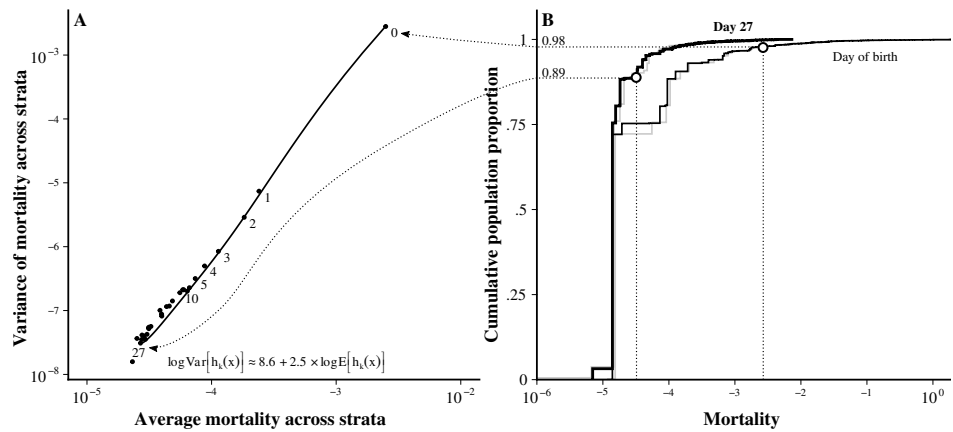

4.2 Shape and level of hazards are heterogeneous

The hazard trajectories predicted by the multilevel Poisson regression model (5) closely match the mortality rate estimates from the life table. They reflect a high degree of heterogeneity among the 252 birth weight×prematurity×Apgar score strata of the U.S. 2009–2012 birth cohort (Figure 3), varying both in level and shape. The heterogeneity is well illustrated by contrasting the lowest-low mortality subgroup (term-born infants with regular birth weight and an Apgar score of ten) with the highest-high mortality stratum (extremely premature infants with extremely low birth weight and an Apgar score of zero). At birth, the force of mortality ranges across five orders of magnitude with a hazard ratio of roughly 244-thousand (!) between the extremes but over time hazards converge as there is a strong positive correlation between the level of the hazard at birth and the rate of mortality decline over the neonatal period. While the lowest-low mortality group features a comparatively flat hazard over age, declining by 39 percent over days 0 to 28, the hazard of the highest-high risk stratum drops by 99.8 percent over the same

period. Hazard trajectories by Apgar score are stratified as expected on a low-high mortality continuum from ten to zero stretching multiple orders of magnitude within all birth weight × prematurity combinations. However, hazards by Apgar score are clustered more closely among extremely preterm infants with extremely low birth-weight compared to term-born infants of regular or high birth weight.

While the hazards are heterogeneous when compared directly, the majority of newborns are born on full-term, with regular birth weight and an Apgar score of nine. Thus the majority of the birth cohort is homogeneous with respect to their hazard trajectory.

Figure 3: Estimated hazard rates versus life table mortality rates over age by prematurity, birth weight, and Apgar score. There is substantial heterogeneity in the level and the shape of the neonatal hazard trajectories ruling out the hypothesis of proportional frailties. The black line in each panel shows the estimated hazard trajectory for the entire birth cohort.
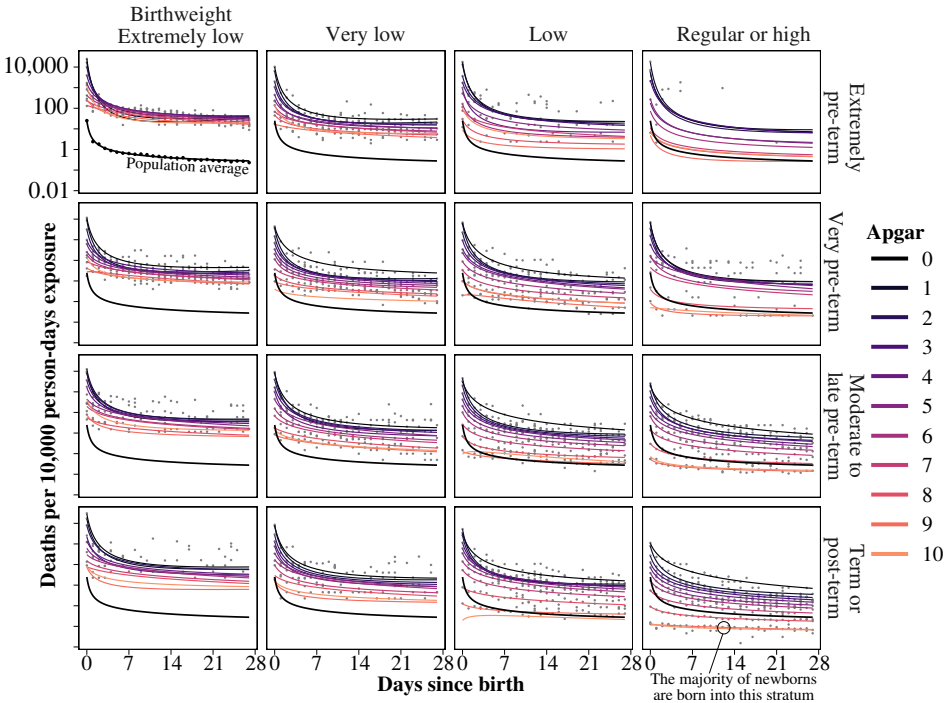

4.3 Mortality selection mainly acts shortly after birth

Population-level mortality during the first day of life is ten times higher than mortality over the following day. This 90 percent mortality decline over 24 hours is mostly the result of a corresponding drop in the stratum specific mortality rates, which, on average, declined by 71 percent between both time points. The remaining 19 percent decline is

explained by a change in population composition due to mortality selection on the day of birth. Over the following time intervals, mortality continues to decline at a fast pace with compositional effects ceasing to contribute substantially to this decline. At no point is the ratio of compositional mortality decline to total mortality decline $\frac{\Delta \overline{m}_j^C}{\Delta \overline{m}_j}$ higher, or even near 0.5 (Table 2).

The results from the discrete-time life table analysis are mirrored by the Vaupel-Zhang decomposition of the hazard's rate of change at different ages. The contribution of mortality rate variance across strata to the slope in the hazard is around 23 percent at birth but merely 4.6 percent 24 hours later staying in the single digits over the remainder of the first month of life (Table 3).

Figure 4: Keeping the population composition fixed at the distribution observed at birth only results in a minor change of the population hazard trajectory.
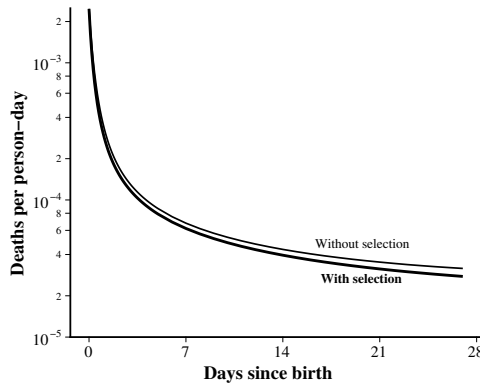
In order to understand the impact of mortality selection on the overall shape of the hazard trajectory I compare the estimated population average hazard over age with a counterfactual population hazard based on the stratum specific hazard estimates of model (5) but assuming the stratum specific population proportions to be fixed, thereby negating the effect of mortality selection on the population hazard. Figure 4 clearly shows that the steep decline in the risk of death following birth is not explained by selection due to differential mortality along the birth weight, Apgar score, and prematurity strata. However, mortality selection has some noticeable effect on the level of risk: Without selection, the hazard of death at day 28 would be an estimated 14 percent higher.

4.4 Stratum specific mortality converges

Following birth, the stratum specific mortality rates rapidly converge, explaining both the sudden decline in variance and mean-mode ratio of the mortality rate distribution (Figures 2, 5).

Table 2: Direct and compositional components of mortality rate decline over the neonatal period.

Age	$\overline{m}_j$		$\Delta\overline{m}_j$	(%)	$\Delta\overline{m}_j^D$	(%)	$\Delta\overline{m}_j^C$	(%)	$\frac{\Delta\overline{m}_j^C}{\Delta\overline{m}_j}$%
0	2.5e-3								
		$\mapsto$	-2.3e-3	-90.4	-1.8e-3	-71.1	-4.8e-4	-19.2	21.3
1	2.4e-4								
		$\mapsto$	-1.8e-4	-73.4	-1.7e-4	-70.1	-7.9e-6	-3.3	4.5
7	6.4e-5								
		$\mapsto$	-2.4e-5	-37.8	-2.3e-5	-36.4	-8.9e-7	-1.4	3.7
14	4.0e-5								
		$\mapsto$	-1.0e-5	-25.9	-1.0e-5	-25.1	-3.4e-7	-0.9	3.3
21	2.9e-5								
		$\mapsto$	-6.3e-6	-21.3	-6.1e-6	-20.9	-1.3e-7	-0.4	2.1
27	2.3e-5								

Age refers to single day age groups.

Table 3: Vaupel-Zhang decomposition of the slope in the hazard of death into direct and compositional components at selected ages.

Age x	$h(x)$	$\dot{\overline{h}}(x)$	$\overline{\dot{h}}(x)$	$\sigma_h^2(x)$	$\nu(x)$%
0	2.5e-3	-1.2e-2	-9.0e-3	2.8e-3	23.5
1	3.1e-4	-3.0e-4	-2.8e-4	1.4e-5	4.6
7	6.2e-5	-6.1e-6	-5.9e-6	1.8e-7	2.9
14	4.0e-5	-1.7e-6	-1.6e-6	6.1e-8	3.6
21	3.1e-5	-7.9e-7	-7.6e-7	3.7e-8	4.7
27	2.8e-5	-4.7e-7	-4.4e-7	3.1e-8	6.5

Figure 5: The ratio of the population average and the modal mortality rates over the first four weeks of life. Following birth population mortality converges towards the low mortality of the most prevalent stratum (Apgar 9, Regular or high birth weight, born term or post-term) which is due to the convergence of high-risk strata.
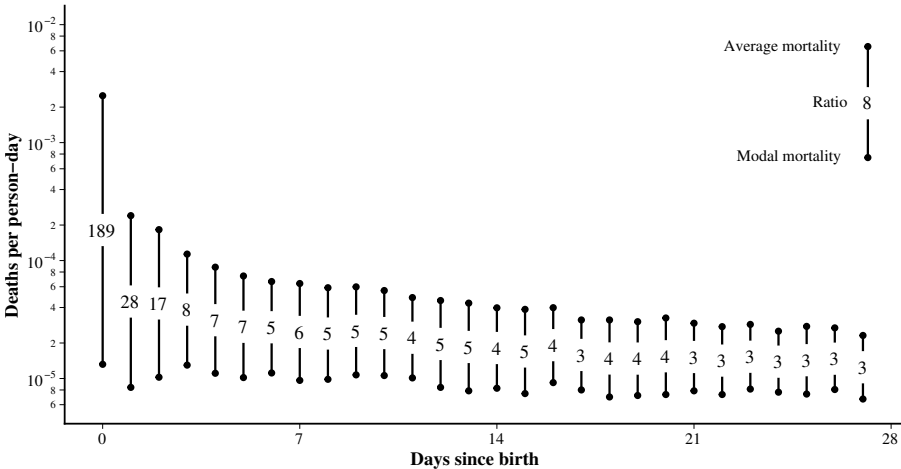
Table 4: Direct and compositional components of mortality rate variance decline over the neonatal period.

Age	v_j		Δv_j	(%)	Δv_j^D	(%)	Δv_j^C	(%)	$\frac{\Delta v_j^C}{\Delta v_j^D}\%$
0	2.5e-3								
		$\mapsto$	-2.3e-3	-90.4	-1.8e-3	-71.1	-4.8e-4	-19.2	21.3
1	2.4e-4								
		$\mapsto$	-1.8e-4	-73.4	-1.7e-4	-70.1	-7.9e-6	-3.3	4.5
7	6.4e-5								
		$\mapsto$	-2.4e-5	-37.8	-2.3e-5	-36.4	-8.9e-7	-1.4	3.7
14	4.0e-5								
		$\mapsto$	-1.0e-5	-25.9	-1.0e-5	-25.1	-3.4e-7	-0.9	3.3
21	2.9e-5								
		$\mapsto$	-6.3e-6	-21.3	-6.1e-6	-20.9	-1.3e-7	-0.4	2.1
27	2.3e-5								

Age refers to single day age groups.

Table 5: Direct and compositional components of mean-mode ratio decline over the neonatal period.

Age	r_j		Δr_j	(%)	Δr_j^D	(%)	Δr_j^C	(%)	$\frac{\Delta r_j^C}{\Delta r_j^D}\%$
0	189								
		$\mapsto$	-160	-84.9	-123	-65.0	-37	-19.9	23.4
1	28								
		$\mapsto$	-21	-76.8	-21	-73.6	0	-3.2	4.2
7	6								
		$\mapsto$	-1	-27.5	-1	-26.0	0	-1.5	5.4
14	4								
		$\mapsto$	-1	-22.1	-1	-21.2	0	-0.9	3.9
21	3								
		$\mapsto$	0	-7.6	0	-7.2	0	-0.5	6.2
27	3								

Age refers to single day age groups.

The variance of death rates over strata is highest at birth and falls by more than 90 percent from day 0 to day 1 of age. The decline in variance is mainly the result of stratum specific death rates converging towards the population average, with only 19 percentage points decline explained by mortality selection. The mortality variance continues to decline substantially over the remainder of the first month of life, with mortality selection never contributing more than 4.5 percent to the total decline (Table 4).

During the first day of life, population mortality is 189 times higher than the mortality of the most prevalent stratum. This ratio declines by 85 percent over the next day of life, with only 20 percentage points of this decline explained by a shifting population composition. During the remainder of the neonatal period, mortality selection only minimally influences the continuing changes in the mean-mode mortality ratio, which eventually arrives at a value of around 3, a drop by 98 percent compared to the value at birth.

5. Discussion

Frailty theory cautions us against taking the population level hazard shape as representative for the risk trajectory of the population members. The analysis of neonatal mortality indeed clearly demonstrates the potential severity of this ecological fallacy, with the majority of newborns experiencing a hazard trajectory radically different from what is observed on the aggregate level. Yet, this difference does not arise from mortality selection but from the non-proportionality of stratum specific hazards. Neonatologists know that the first 24 hours of life are the most critical in determining the further survival of the preterm infant. Once the one-day threshold is passed, the survival chances not only drastically improve but, as shown in this paper, converge towards the survival chances of healthy term-born infants. The popular proportional frailty model is severely misspecified for situations where an individual may be saved due to medical intervention, i.e., where the physiological state of an individual and thus the associated frailty changes abruptly. In consequence, even if the model fits the aggregate age pattern well, as it does in case of the Hougaard-Gompertz fit to the age trajectory of infant mortality (Figure 1), inferences drawn from it are biased.

To escape the strong assumptions of frailty models, I based the analysis in this paper on the *observed* heterogeneity in age-specific mortality across mutually exclusive population strata but, in turn, left open the possibility for *hidden* heterogeneity within a particular stratum. A further division of the population along the presence and severity of congenital malformations may uncover that the mortality decline following birth is primarily the effect of the vanishing subpopulation of newborns featuring congenital disabilities incompatible with life. While this particular hypothesis can be tested in a future publication, one can never exclude the possibility of further unobserved quantities influencing observed outcomes. All that can be confidently said based on the data analyzed above is that the heterogeneity in risk along birth weight, Apgar score, and prematurity has only a minimal impact on the shape of the population level age trajectory of neonatal mortality.

A peculiar finding is the near-perfect power-law relationship between the mean and the variance of the distribution of stratum-specific mortality rates over age – a further example of "Taylor's law" in demography: Analyzing data on the spatial distribution of various species Taylor (1961) observed that the variance in population count per unit area is well predicted by a power of the average population count. Similar mean-variance relationships have been found for population densities in human populations (Cohen et al. 2013; Naccarato and Benassi 2018) and recently for time-series of age-specific mortality rates and rates of mortality improvement (Bohk et al. 2015; Cohen et al. 2018). Whereas the studies by Cohen, Bohk, and Rau are based on the concept of a "temporal mean/variance," i.e., mean and variance are estimated from observations repeated over age or period, I contribute the first demonstration of Taylor's law in human mortality based on the distribution of risk at any single point in time within a cohort of individuals followed over age.

6. Conclusion

The sudden drop in population mortality following birth is predominantly the result of stratum specific hazard rates quickly converging towards rather low levels. The compositional shift in the population resulting from mortality selection only has a minimal impact on the age trajectory of neonatal mortality.

It remains to be seen if the simple power-law expression that describes the time-evolution of mean and variance in mortality among a cohort of newborns replicates in other populations and at other ages. The prospect of having a 2-parameter expression for the mortality dynamics in a heterogeneous aging cohort may motivate future inquiries in that direction.

A. The Hougaard-Gompertz frailty model

Let the hazard of death at age x for a newborn with frailty z be $h_z(x|z) = zh_0(x)$ with z drawn from the three parameter frailty distribution proposed by Hougaard (1986) denoted with $Z \sim \text{Houg}(x; \alpha, \delta, \theta)$ and Gompertz baseline hazard $h_0(x) = ae^{bx}$ with corresponding cumulative hazard $H_0(x) = \int_0^x h_0(s)\,ds = a(e^{bx} - 1)/b$. The average frailty among the survivors at x then is $\text{E}[Z|X \geq x] = \delta\,(\theta + H_0(x))^{\alpha-1}$ (Hougaard 1986, p. 393). Substituting $\delta = \frac{1}{\theta^{\alpha-1}}$ eliminates one parameter from the model and fixes the average frailty at birth at one. The population hazard then is

$$h(x) = \text{E}[Z|X = x]h_0(x) = \left(\frac{\theta + a(e^{bx} - 1)/b}{\theta}\right)^{\alpha-1} ae^{bx},$$

with log-hazard

$$\log h(x) = (\alpha - 1)\log\left(\frac{\theta + a(e^{bx} - 1)/b}{\theta}\right) + \log a + bx$$

The above hazard is fitted to daily death counts D_j and exposure times E_j of the U.S. 2009–2012 infant life table via a Bayesian non-linear Poisson regression model

$$D_j \sim \text{Pois}(\lambda_j E_j) \text{ for } j = 1, \ldots, 365$$
$$\lambda_j = e^{\log h(x_j; a, b, \theta, \alpha)}$$

with parameter transformations

$$a = e^{\beta_0}$$
$$b = \beta_1$$
$$\theta = e^{\beta_2}$$
$$\alpha = \log\frac{\beta_3}{1 - \beta_3},$$

and priors

$$\beta_0 \sim \text{N}(-8, 1)$$
$$\beta_1 \sim \text{N}(-0.03, 1)$$
$$\beta_2 \sim \text{N}(-6, 1)$$
$$\beta_3 \sim \text{N}(0.5, 1).$$

B. The quadratic hazard over log age

Quadratic relationships between log age and log mortality can be captured by the expression

$$h(x) = e^{\beta_0 + \beta_1 \log(x+1) + \beta_2 \log^2(x+1)},$$

with derivative wrt. x

$$h'(x) = (x+1)^{\beta_1 - 1} e^{\beta_0 + \beta_2 \log^2(x+1)} (\beta_1 + 2\beta_2 \log(x+1)),$$

and survival function

$$S(x) = \exp\left(\frac{\sqrt{\pi} \exp\left(b_0 - \frac{(b_1+1)^2}{4b_2} \right) \left(\operatorname{erfi}\left(\frac{b_1+1}{2\sqrt{b_2}} \right) - \operatorname{erfi}\left(\frac{b_1 + 2b_2 \log(x+1) + 1}{2\sqrt{b_2}} \right) \right)}{2\sqrt{b_2}} \right).$$

An efficient algorithm for the evaluation of the imaginary error function $\operatorname{erfi}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(t^2 \right) \, dt$ exists (Poppe and Wijers 1990) and is implemented in the C++ library Faddeeva which can be called from within R via the package RcppFaddeeva.

C. The Vaupel-Zhang equality

I find it instructive to derive the Vaupel-Zhang equality from a simple finite mixture distribution of survival times as it gives justification to the weighted averages employed in this paper. Once the marginal hazard rate has been shown to be a weighted average of hazards among the survivors the proof is the same as Vaupel and Zhang (2010).

Let X be the positive real-valued random variable "age at death" and K be a random index denoting membership to the kth population stratum. The marginal density of ages at death is given by the discrete mixture distribution

$$f(x) = \sum_k f(x|k)\mathrm{P}(k),$$

where $f(x|k)$ is the conditional density of deaths and $\mathrm{P}(k)$ the probability of membership in stratum k. By definition the conditional and marginal survival functions are

$$S(x|k) = \mathrm{P}(X > x|k) = \int_x^\infty f(x|k)\,\mathrm{d}x$$

$$S(x) = \mathrm{P}(X > x) = \int_x^\infty f(x)\,\mathrm{d}x,$$

which give rise to the conditional and marginal hazards via the relationships

$$h(x|k) = -S'(x|k)/S(x|k)$$

$$h(x) = -S'(x)/S(x),$$

where the prime mark denotes the derivative with respect to age x.

The Vaupel-Zhang equality states that for the survivors at age x the age-derivative of their average hazard is equal to the average age derivative minus the variance of the hazards, in their notation

$$\dot{\bar{h}}(x) = \overline{h}(x) - \sigma_h^2(x),$$

where

$$\dot{\bar{h}}(x) = \left[\sum_k h(x|k)\mathrm{P}(k|X > x)\right]'$$

$$\overline{h}(x) = \sum_k h'(x|k)\mathrm{P}(k|X > x)$$

$$\sigma_h^2(x) = \sum_k h(x|k)^2 \mathrm{P}(k|X > x) - \left[\sum_k h(x|k)\mathrm{P}(k|X > x)\right]^2.$$

Proof:

Bayes Theorem gives the probability of being in stratum k given survival to age x as

$$P(k|X > x) = \frac{P(X > x|k)P(k)}{P(X > x)} = \frac{S(x|k)P(k)}{S(x)}$$

The marginal survival $S(x)$ can be calculated from conditional survival $S(x|k)$ and stratum distribution $P(k)$ via the law of total probability as

$$S(x) = P(X > x)$$
$$= \sum_k P(X > x|k)P(k)$$
$$= \sum_k S(x|k)P(k),$$

with age derivative

$$S'(x) = \sum_k S'(x|k)P(k)$$
$$= \sum_k -h(x|k)S(x|k)P(k).$$

By substituting $S'(x)$ into $h(x) = -\frac{S'(x)}{S(x)}$ one can express the marginal hazard as a weighted average of the hazards of the survivors:

$$h(x) = -\frac{\sum_k -h(x|k)S(x|k)P(k)}{S(x)}$$
$$= \sum_k \frac{h(x|k)S(x|k)P(k)}{S(x)}$$
$$= \sum_k h(x|k)P(k|X > x).$$

Taking the derivative of $h(x)$ gives

$$\dot{h}(x) = \sum_k h'(x|k)P(k|X > x) + \sum_k h(x|k)P'(k|X > x)$$
$$= \bar{\dot{h}}(x) + \sum_k h(x|k)P'(k|X > x).$$

The second term can be manipulated to yield

$$\sum_k h(x|k)\mathrm{P}'(k|X > x) = \sum_k h(x|k)\left(\frac{\mathrm{P}(k)S'(x|k)}{S(x)} - \frac{\mathrm{P}(k)S'(x)S(x|k)}{S(x)^2}\right)$$

$$= \sum_k \frac{h(x|k)\mathrm{P}(k)S'(x|k)}{S(x)} - \sum_k \frac{h(x|k)\mathrm{P}(k)S'(x)S(x|k)}{S(x)^2}$$

$$= \sum_k \frac{h(x|k)\mathrm{P}(k)S'(x|k)}{S(x)} - \frac{S'(x)}{S(x)} \sum_k h(x|k)\mathrm{P}(k|X > x)$$

$$= \sum_k \frac{h(x|k)\mathrm{P}(k)S'(x|k)}{S(x)} - h(x)^2$$

$$= \sum_k \frac{-h(x|k)^2\mathrm{P}(k)S(x|k)}{S(x)} - h(x)^2$$

$$= -\sum_k h(x|k)^2\mathrm{P}(k|X > x) - h(x)^2$$

$$= -\sigma_h^2(x),$$

completing the proof.

References

V. Apgar. A proposal for a new method of evaluation of the newborn infant. *Current Researches in Anesthesia and Analgesia*, 32(4):260–267, 1953. doi:10.1213/ANE.0b013e31829bdc5c.

D. Avraam, S. Arnold, D. Jones, and B. Vasiev. Time-evolution of age-dependent mortality patterns in mathematical model of heterogeneous human population. *Experimental Gerontology*, 60:18–30, 2014. doi:10.1016/j.exger.2014.09.006.

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi:10.18637/jss.v067.i01.

R. E. Beard. Appendix: Note on some mathematical mortality models. In G. E. W. Wolstenholme and M. O'Connor, editors, *The Lifespan of Animals. Ciba Foundation Colloquium on Ageing*, pages 802–811, Boston, 1959. Little, Brown. doi:10.1002/9780470715253.app1.

C. Bohk, R. Rau, and J. E. Cohen. Taylor's power law in human mortality. *Demographic Research*, 33(21):589–610, 2015. doi:10.4054/DemRes.2015.33.21.

G. Boole. *Calculus of finite differences*. Chelsea Publishing Company, New York, 4th edition, 1880.

B. M. Casey, D. D. McIntire, and K. J. Leveno. The continuing value of the apgar score for the assessment of newborn infants. *New England Journal of Medicine*, 344(7): 467–471, 2001. doi:10.1056/NEJM200102153440701.

J. E. Cohen, M. Xu, and H. Brunborg. Taylor's law applies to spatial variation in a human population. *Genus*, 69(1):25–60, 2013.

J. E. Cohen, C. Bohk-Ewald, and R. Rau. Gompertz, Makeham, and Siler models explain Taylor's law in human mortality data. *Demographic Research*, 38(29):773–841, 2018. doi:10.4054/DemRes.2018.38.29.

F. Colchero and B. Y. Kiyakoglu. Beyond the proportional frailty model: Bayesian estimation of individual heterogeneity on mortality parameters. *Biometrical Journal*, pages 1–12, 2019. doi:10.1002/bimj.201800280.

A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press, Cambridge, 2007. ISBN 978-0-511-26878-6. doi:10.1017/cbo9780511790942.

J. M. Hoem. Identifiability in hazard models with unobserved heterogeneity: The compatibility of two apparently contradictory results. *Theoretical Population Biology*, 37 (1):124–128, 1990. doi:10.1016/0040-5809(90)90030-y.

P. Hougaard. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*, 71(1):75–83, 1984. doi:10.1093/biomet/71.1.75.

P. Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396, 1986. doi:10.1093/biomet/73.2.387.

J. J. Hsieh. Construction of expanded infant life tables: A method based on a new mortality law. *Mathematical Biosciences*, 76(2):221–242, 1985. doi:10.1016/0025-5564(85)90006-9.

E. M. Kitagawa. Components of a difference between two rates. *Journal of the American Statistical Association*, 50(272):1168–1194, 1955. doi:10.1080/01621459.1955.10501299.

D. A. Levitis. Before senescence: the evolutionary demography of ontogenesis. *Proceedings of the Royal Society B*, 278(1707):801–809, 2011. doi:10.1098/rspb.2010.2190.

D. A. Levitis and D. E. Martínez. The two halves of U-shaped mortality. *Frontiers in Genetics*, 4(31):1–6, 2013. doi:10.3389/fgene.2013.00031.

T. I. Missov and J. W. Vaupel. Mortality implications of mortality plateaus. *SIAM Review*, 57(1):61–70, 2015. doi:10.1137/130912992.

A. Naccarato and F. Benassi. On the relationship between mean and variance of world's human population density: A study using Taylor's power law. *Letters in Spatial and Resource Sciences*, 11(3):307–314, 2018. doi:10.1007/s12076-018-0214-5.

National Center for Health Statistics. Birth cohort linked birth-infant death data files (U.S. data), 2016. URL ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/DVS/cohortlinkedus/.

J. H. Park, Y. S. Chang, S. Y. Ahn, S. I. Sung, and W. S. Park. Predicting mortality in extremely low birth weight infants: Comparison between gestational age, birth weight, Apgar score, CRIB II score, initial and lowest serum albumin levels. *PLoS ONE*, 13 (2):1–10, 2018. doi:10.1371/journal.pone.0192232.

M. M. Pollack, M. A. Koch, D. A. Bartel, I. Rapoport, R. Dhanireddy, A. A. E. El-Mohandes, K. Harkavy, and K. N. S. Subramanian. A comparison of neonatal mortality risk prediction models in very low birth weight infants. *Pediatrics*, 105(5):1051–1057, 2000. doi:10.1542/peds.105.5.1051.

G. P. M. Poppe and C. M. J. Wijers. More efficient computation of the complex error function. *ACM Transactions on Mathematical Software (TOMS)*, 16(1):38–46, 1990.

A. Remund. *Jeunesses vulnérables? Mesures, composantes et causes de la surmortalité des jeunes adultes*. PhD thesis, Université de Genève, 2015.

D. R. Steinsaltz and K. W. Wachter. Understanding mortality rate deceleration and heterogeneity. *Mathematical Population Studies*, 13(1):19–37, 2006. doi:10.1080/08898480500452117.

L. R. Taylor. Aggregation, variance and the mean. *Nature*, 189(4766):732–735, 1961. doi:doi.org/10.1038/2F189732a0.

J. Trussell and T. Richards. Correcting for unmeasured heterogeneity in hazard models using the heckman-singer procedure. *Sociological Methodology*, 15:242–276, 1985. doi:10.2307/270852.

J. W. Vaupel and J. R. Carey. Compositional interpretations of medfly mortality. *Science*, 260(5114):1666–1667, 1993. doi:10.1126/science.8503016.

J. W. Vaupel and T. I. Missov. Unobserved population heterogeneity: A review of formal relationships. *Demographic Research*, 31(1):659–686, 2014. doi:10.4054/DemRes.2014.31.22.

J. W. Vaupel and A. I. Yashin. The deviant dynamics of death in heterogeneous populations. Technical report, Laxenburg, Austria, 1983. URL http://user.demogr.mpg.de/jwv/pdf/IIASA-83-001.pdf.

J. W. Vaupel and A. I. Yashin. Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *The American Statistician*, 39(3):176–185, 1985. doi:10.2307/2683925.

J. W. Vaupel and Z. Zhang. Attrition in heterogeneous cohorts. *Demographic Research*, 23(26):737–748, 2010. doi:10.4054/DemRes.2010.23.26.

J. W. Vaupel, K. G. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–54, 1979. doi:10.2307/2061224.

A. Wienke. *Frailty Models in Survival Analysis*. Biostatistics Series. Chapman and Hall, Boca Raton, 2011. ISBN 978-1-4200-7388-1. doi:10.1111/j.1541-0420.2012.01769.x.

A. I. Yashin, I. A. Iachine, and A. S. Begun. Mortality modeling: A review. *Mathematical Population Studies*, 8(4):305–332, 2000. doi:10.1080/08898480009525489.