How Do Populations Aggregate?

Dennis Feehan
University of California, Berkeley

Elizabeth Wrigley-Field†
University of Minnesota

†Address correspondence to Elizabeth Wrigley-Field, University of Minnesota, Department of Sociology, 1156 Social Science Building, 267 19th Ave S., Minneapolis, MN 55455 (email: ewf@umn.edu). Support for this work was provided by the Berkeley Population Center (P2C HD073964) and the Minnesota Population Center (P2C HD041023).

How do populations aggregate? *

Dennis M. Feehan† and Elizabeth Wrigley-Field‡

Abstract

BACKGROUND

Understanding the relationship between populations at different scales plays an important role in many demographic analyses.

OBJECTIVE

We show that when a population can be partitioned into subgroups, the death rate for the entire population can be written as the weighted harmonic mean of the death rates in each subgroup, where the weights are given by the numbers of deaths in each subgroup. This decomposition can be generalized to other types of occurrence-exposure rate. Using different weights, the death rate for the entire population can also be expressed as an arithmetic mean of the death rates in each subgroup.

CONTRIBUTION

We use these relationships as a starting point for investigating how demographers can correctly aggregate rates across non-overlapping subgroups. Our analysis reveals conceptual links between classic demographic models and length-biased sampling. To illustrate how the harmonic mean can suggest new interpretations of demographic relationships, we present as an application a new expression for the frailty of the dying, given a standard demographic frailty model.

---

†UC Berkeley, feehan@berkeley.edu
‡University of Minnesota, ewf@umn.edu

1 Relationship

Given a population and a partition of the population into subgroups, what is the relationship between the death rate in the subgroups and the death rate for the aggregate population? We show that occurrence-exposure rates can be understood to aggregate across scales according to an elegant and well-understood mathematical relationship: the weighted harmonic mean, which is the inverse of a weighted integral or sum of inverses.

Definition. Let $f : [\alpha, \beta] \rightarrow \mathbb{R}$ be a continuous, positive function $f(x) > 0$ and let $w : [\alpha, \beta] \rightarrow \mathbb{R}$ be a continuous non-negative function $w(x) \geq 0$. Then the **Weighted Harmonic Mean** of f with weights given by w is

$$A_H[f(x); w(x)] = \frac{\int_\alpha^\beta w(x)dx}{\int_\alpha^\beta \frac{w(x)}{f(x)}dx} = \left(\frac{\int_\alpha^\beta w(x)(f(x))^{-1}dx}{\int_\alpha^\beta w(x)dx} \right)^{-1}. \tag{1}$$

Result. *Suppose that a population is a mixture of people with different values of some continuous trait $u \in [\alpha, \beta]$. Let $d(a, u)$ be the number of deaths at exact age a to people with mixing trait value u, and let $\mu(a, u)$ be the hazard faced at exact age a by people with trait value u. Assume $d(a, u)$ and $\mu(a, u)$ are positive and continuous when viewed as functions of u. Then the aggregate hazard at age a, $\bar{\mu}(a)$, is the weighted harmonic mean of $\mu(a, u)$, with weights given by the number of deaths, $d(a, u)$:*

$$\bar{\mu}(a) = A_H[\mu(a, u); d(a, u)] = \frac{\int_\alpha^\beta d(a, u)du}{\int_\alpha^\beta \frac{d(a,u)}{\mu(a,u)}du} = \left(\frac{\int_\alpha^\beta d(a, u)(\mu(a, u))^{-1}du}{\int_\alpha^\beta d(a, u)du} \right)^{-1}. \tag{2}$$

The relationship holds whenever a population can be partitioned into subgroups – i.e., when the population can be divided into a set of mutually exclusive and collectively exhaustive subgroups. There are many potentially interesting ways that populations can be understood to be continuous mixtures of non-overlapping subpopulations: for example, at any moment, a population is a mixture of people with different blood pressure, cholesterol, BMI, height, and so forth. The decomposition does not depend upon any particular feature of death rates; it will also hold for any other type of occurrence-exposure rate.

2 Proof

Proof. Let $\ell(a,u)$ be the number of survivors to exact age a with trait u, where the values of u partition the population. Then

$$\bar{\mu}(a) = \frac{\int_\alpha^\beta d(a,u)du}{\int_\alpha^\beta \ell(a,u)du} = \frac{\int_\alpha^\beta d(a,u)du}{\int_\alpha^\beta \frac{\mu(a,u)\ell(a,u)}{\mu(a,u)}du} = \frac{\int_\alpha^\beta d(a,u)du}{\int_\alpha^\beta \frac{d(a,u)}{\mu(a,u)}du}. \tag{3}$$

$\square$

3 Related relationships

Discrete version

A harmonic mean can also be defined for discrete quantities, which leads to an analogous result for a population that has been partitioned into a discrete set of subgroups. Suppose that a population has been partitioned into a countable set of subgroups, indexed by i, and let the death count, exposure, and death rate in subgroup i between ages a and $a+n$ be denoted $_nD_a^i$, $_nL_a^i$, and $_nM_a^i = \frac{_nD_a^i}{_nL_a^i} > 0$, respectively. Then

$$_nM_a = \frac{\sum_i {}_nD_a^i}{\sum_i {}_nL_a^i} = \frac{\sum_i {}_nD_a^i}{\sum_i \frac{_nL_a^i {}_nM_a^i}{_nM_a^i}} = \frac{\sum_i {}_nD_a^i}{\sum_i \frac{_nD_a^i}{_nM_a^i}}, \tag{4}$$

where $_nM_a$ is the aggregate death rate between ages a and $a+n$. In words, the aggregate death rate is the discrete harmonic mean of the subgroup death rates, with weights given by the number of deaths in each subgroup.

Length-biased sampling

The harmonic mean often arises in applications of *length-biased* sampling (see, e.g., de Carvalho (2016) and Patil (2014) for reviews). Under length-biased sampling, the probability of observing a characteristic is proportional to the value of that characteristic.

Formally, suppose that the population-level distribution of some characteristic $x > 0$ is given by a probability density function $f(x)$. (We use x to denote the general case of any single characteristic, as opposed to the earlier example of a population stratified simultaneously by age a and some non-age trait u.) Now consider $y > 0$, a length-biased observation from f. Then the probability density function describing the distribution of the length-biased observation y is given by $f^*(y)$:

$$f^*(y) = \frac{y f(y)}{E_f[x]}, \tag{5}$$

where $E_f[x] = \int_0^\infty x f(x) dx$ is the usual arithmetic mean of x taken with respect to the non-length biased distribution f. From this observation mechanism, de Carvalho (2016) shows that

$$
\begin{aligned}
E_{f^*}\left(\frac{1}{y}\right) &= \int_0^\infty \frac{1}{u} \times \frac{u f(u)}{E_f[x]} du \\
&= \frac{1}{E_f[x]} \int_0^\infty f(u) du \\
&= \frac{1}{E_f[x]},
\end{aligned} \tag{6}
$$

where E_{f^*} is the expectation taken with respect to the length-biased distribution $f^\star$. Thus, Equation 6 shows that $E_f[x] = \frac{1}{E_{f^*}[1/y]}$. Since the harmonic mean of the length-biased distribution is $A_H[y, f^\star] = \frac{\int_0^\infty f^\star(u) du}{\int_0^\infty \frac{1}{u} f^\star(u) du} = \frac{1}{E_{f^*}[1/y]}$, the relationship in Equation 6 says that the arithmetic mean of the non length-biased distribution is equal to the harmonic mean of the length-biased distribution, $E_f[x] = \frac{1}{E_{f^*}[1/y]}$. Thus, when we wish to recover the population average from length-biased samples, we should use the harmonic mean.

What happens if, instead, we take the arithmetic mean of length-biased samples? Recall that, when $y > 0$,

$$
\begin{aligned}
E_{f^*}[y] &= \int_0^\infty u \times \frac{u f(u)}{E_f[x]} du \\
&= \frac{1}{E_f[x]} \int_0^\infty u^2 f(u) du \\
&= \frac{E_f[x^2]}{E_f[x]}.
\end{aligned} \tag{7}
$$

Equation 7 is the second moment of the non-length biased distribution, $E_f[x^2]$, divided by the first moment of the non-length biased distribution, $E_f[x]$. We can use the definition of variance to help make Equation 7 more interpretable: by definition, $\text{var}_f[x] = E_f[x^2] - E_f[x]^2$; so, rearranging, we have $E_f[x^2] = \text{var}_f[x] + E_f[x]^2$. Plugging this relationship in Equation 7, we obtain

$$
\begin{aligned}
E_{f^*}[y] &= \frac{1}{E_f[x]} \left[E_f[x]^2 + \text{var}_f[x] \right] \\
&= E_f[x] \left[1 + \text{cv}_f^2[x] \right],
\end{aligned}
\tag{8}
$$

where $\text{cv}_f^2[x] = \frac{\text{var}_f[x]}{E_f^2[x]}$ is the squared coefficient of variation of x taken with respect to the distribution f (Sen, 1987). In words, Equation 8 shows that the arithmetic mean of length-biased samples will differ from the underlying population mean by a factor that increases in the squared coefficient of variation of samples from the non-length biased distribution.

In the case of death rates, Equation 7 can be used to show that the arithmetic mean of subpopulation mortality weighted by the number of deaths is affected by length bias:

$$
\begin{aligned}
E_{d(x)}[\mu(x)] &= \frac{\int_0^\infty \mu(x)d(x)dx}{\int_0^\infty d(x)dx} \\
&= \frac{\int_0^\infty \mu(x)^2\ell(x)dx}{\int_0^\infty \mu(x)\ell(x)dx} \\
&= \frac{E_{\ell(x)}[\mu(x)^2]}{E_{\ell(x)}[\mu(x)]},
\end{aligned}
\tag{9}
$$

where $E_{d(x)}$ is the expectation taken with respect to the distribution of deaths at each age, $d(x)$, and $E_{\ell(x)}$ is the expectation taken with respect to the distribution of survivors at each age, $\ell(x)$.

Equation 9 has the form of an the arithmetic mean of length-biased samples (Equation 7). We illustrate these relationships in the appendix with an illustration based on aggregating subnational life tables. We also show a different example of how this length-biasing relationship arises in demography in Section 5.

Two ways of aggregating rates

Closer inspection of Equation 4 reveals that the arithmetic mean can also be used to aggregate across subgroups, now using the exposure as the weights, rather than the deaths. So, using $A_A[X; W]$ to refer to the arithmetic mean of X with weights given by W, we have $A_H[M; D] = A_A[M; L]$, or

$$\underbrace{\frac{{}_nD_a}{{}_nL_a}}_{\text{Aggregate death rate}} = \underbrace{\frac{\sum_i {}_nD_a^i}{\sum_i \frac{{}_nD_a^i}{{}_nM_a^i}}}_{\substack{\text{Harmonic mean} \\ \text{weighted by} \\ \text{deaths}}} = \underbrace{\frac{\sum_i {}_nM_a^i \, {}_nL_a^i}{\sum_i {}_nL_a^i}}_{\substack{\text{Arithmetic mean} \\ \text{weighted by} \\ \text{exposure}}}. \tag{10}$$

The two decompositions in Equation 10 distinguish between weighting subgroup death rates by exposure (arithmetic mean) or by the number of deaths (harmonic mean). The conceptual key is that the harmonic mean is a natural average to use for quantities whose reciprocals aggregate according to the usual (arithmetic) mean. This explains why the harmonic mean is relevant in length-biased sampling: Equation 6 shows how taking the reciprocal 'adjusts' for the length bias, because $\frac{1}{u} u f(u) = f(u)$. (We discuss this in greater detail in below.)

For death rates, the arithmetic mean decomposition in Equation 10 says that, fixing exposure, deaths will be observed in direct proportion to the subgroups' death rates. For example, suppose that subgroup 1 has twice the death rate of subgroup 2, (i.e., $M^1 = 2M^2$), and that we observe a fixed amount of exposure, say L, from each subgroup. Then subgroup 1 will be expected to contribute twice as many deaths to the aggregate as subgroup 2 (i.e., $D^1 = M^1 \, L = 2M^2L = 2D^2$). Thus, when weighted by exposure, the death rates aggregate arithmetically.

However, the harmonic mean decomposition in Equation 10 says that, fixing deaths, exposure will be not be observed in proportion to subgroups' death rates; instead, exposure will be *inversely* proportional to subgroups' death rates. For example, suppose again that subgroup 1 has twice the death rate of subgroup 2, (i.e., $M^1 = 2M^2$), but we now observe a fixed number of deaths, say D, in each subgroup. If subgroup 1 has twice the death rate of subgroup 2, then we expect that subgroup 1 will contribute half as much exposure as subgroup 2 to the aggregate (i.e., $L^1 = \frac{D}{M^1} = \frac{D}{2M^2} = \frac{1}{2}L^2$). Thus, each subgroup's contribution to the aggregate exposure is proportional to the reciprocal of its death rate, and the death rates aggregate according to the harmonic mean. (Note that this also implies that the expected waiting time to each death – i.e., the reciprocal of the death rate – aggregates in the usual way, according

to the arithmetic mean.)

4 History

There is statistical literature on the properties and uses of the harmonic mean (e.g., de Carvalho, 2016; Sen, 1987). Demographers originally discussed the harmonic mean explicitly in the context of selection problems and heterogeneity in fertility (Sheps, 1964; Sheps, Menken, and Radick, 1973) and mortality (Keyfitz and Littman, 1979), or implicitly in the context of understanding family size, viewed from the perspective of mothers and children (Preston, 1976). However, after these papers, demographers have made relatively little use of the harmonic mean, in spite of the field's focus on aggregate rates and their relationship to underlying individual rates. (An exception is Schoen (2013), who investigated the harmonic mean as it relates to two-sex population models.)

5 Application

Aggregation is a central focus of the literature on heterogeneity in mortality (e.g., Vaupel, Manton, and Stallard (1979); Vaupel and Missov (2014)). We illustrate the harmonic mean by applying it to derive a new formula for a classic result in that literature.

In what Vaupel and Missov (2014) calls the 'relative risks and fixed frailty' model, heterogeneity in death rates is captured by associating each individual in the population with a frailty parameter $z > 0$. People with frailty parameter z face hazards that are related to a baseline via

$$\mu(a, z) = z\mu_s(a), \tag{11}$$

where $\mu_s(a) \equiv \mu(a, 1)$ is the hazard faced by a 'standard individual', with frailty $z = 1$, at age a[1]. Note that this model describes a situation in which our result applies: the population can be understood to be a continuous mixture of people who have different frailty parameters.

We will now see that, at age a, the frailties of people who die are length-biased samples of the

---

[1]To be consistent, we refer to the age variable as a here; note that Vaupel and Missov (2014) use x for age in their paper.

frailties of people who are alive. Following Vaupel and Missov (2014, p. 663), let $s(a, z)$ be the survivorship to age a among those with frailty parameter z and let $\bar{s}(a)$ be the aggregate survivorship to age a. Let $\pi(a, z)$ be the density of people with frailty parameter z among survivors to age a, and let $\bar{z}(a)$ be the average frailty of survivors at age a. Finally, let $d(a, z)$ be the density of deaths at age a to people with frailty parameter z and let $d(a)$ be the density of deaths at age a. Then the probability that a death aged a has frailty z is given by:

$$
\begin{aligned}
\frac{d(a, z)}{d(a)} &= \frac{\mu(a, z)s(a, z)\pi(0, z)}{\int_0^\infty \mu(a, z)s(a, z)\pi(0, z)dz} \\
&= \frac{\mu(a, z)\pi(a, z)\bar{s}(a)}{\int_0^\infty \mu(a, z)\pi(a, z)\bar{s}(a)dz} \\
&= \frac{\mu(a, z)\pi(a, z)}{\int_0^\infty \mu(a, z)\pi(a, z)dz} \\
&= \frac{z\mu_s(a)\pi(a, z)}{\int_0^\infty z\mu_s(a)\pi(a, z)dz} \\
&= \frac{z\,\pi(a, z)}{\int_0^\infty z\,\pi(a, z)dz} \\
&= \frac{z\,\pi(a, z)}{\bar{z}(a)},
\end{aligned}
\tag{12}
$$

where the second line uses the fact that $s(a, z)\pi(0, z) = \pi(a, z)\bar{s}(a)$ (Vaupel and Missov, 2014, p. 663). Equation 12 shows that the probability that a death aged a has frailty z is proportional to z times the density of frailties among survivors aged a, divided by the average frailty among survivors aged a. Thus, Equation 12 has exactly the same form as a length-biased sampling density (Equation 5). In words, the frailty of someone who dies at age a can be understood as a length-biased sample from the distribution of frailties among people who survive to age a.

As a length-biased sample, two results follow immediately. First, Equation 8 showed (as is well-known) that the arithmetic mean of length-biased samples can be written in terms of the mean and squared coefficient of variation in the non-length-biased population distribution; in the relative risks and fixed frailty model, Equation 8 provides an expression for $z^\dagger(a)$, the average frailty of the dead:

$$
z^\dagger(a) = \bar{z}(a)\left[1 + \mathrm{cv}_z^2(a)\right],
\tag{13}
$$

where $\mathrm{cv}_z^2(a)$ is the coefficient of variation in the frailty parameter of people at age a. Equation

13 has been previously derived by Vaupel, Manton, and Stallard (1979, p. 442); a proof is also briefly discussed in Vaupel and Missov (2014, p. 667–668).

Second, Equation 6 showed that the harmonic mean of length-biased samples will recover the population (non-length-biased) mean. In the relative risks and fixed frailty model, Equation 6 provides an expression for the average frailty of people who survive to age a, in terms of the frailties of people who die at age a:

$$
\begin{aligned}
A_H[z, d(a, z)] &= \frac{\int_0^\infty d(a, z)dz}{\int_0^\infty \frac{d(a,z)}{z}dz} \\
&= \frac{\int_0^\infty \mu(a, z)\pi(a, z)dz}{\int_0^\infty \frac{\mu(a,z)\pi(a,z)}{z}dz} \\
&= \frac{\int_0^\infty z\mu_s(a)\pi(a, z)dz}{\int_0^\infty \frac{z\mu_s(a)\pi(a,z)}{z}dz} \\
&= \frac{\int_0^\infty z\,\pi(a, z)dz}{\int_0^\infty \pi(a, z)dz} \\
&= \int_0^\infty z\,\pi(a, z)dz \\
&= \bar{z}(a).
\end{aligned}
\tag{14}
$$

Equation 14 reveals that, under the relative risks and fixed frailty model, the average frailty of people who survive to age a is equal to the harmonic mean of the frailties of people who die at age a. So, if the relative risks and fixed frailty model held in a real population, and z could be measured or estimated only from a sample of deaths at age a, then Equation 14 says that we could still recover the population average frailty by taking the harmonic mean of the frailties among the deaths. (Similarly, any way of estimating the frailty of the individuals alive at some moment would also produce an estimate of the frailty of those dying at the same moment.)

Finally, we note that the relationship between the harmonic mean and reciprocals suggests a new interpretation of the frailty parameter, z, by focusing attention on the reciprocal of frailty, $\frac{1}{z}$. This frailty reciprocal, $\frac{1}{z}$, represents the number of deaths to 'standard individuals' that are expected before a death to a single individual of frailty z occurs. As such, it can be conceptualized as a form of waiting time, or exposure, needed to produce a death at frailty z, expressed in a relative scale.

As this example illustrates, bringing harmonic means more widely into demographic analysis

can create analogies and conceptual linkages between demographic quantities that are not otherwise obvious. This is particularly true for questions of aggregation, in which a length-biased selection process is often at play. More generally, the harmonic and arithmetic mean relationships discussed above illustrate how demographers can correctly aggregate rates across non-overlapping subgroups.

References

de Carvalho, M. (2016). Mean, What do You Mean? *The American Statistician* 70(3): 270–274. URL https://doi.org/10.1080/00031305.2016.1148632.

Keyfitz, N. and Littman, G. (1979). Mortality in a heterogeneous population. *Population studies* 33(2): 333–342.

Patil, G.P. (2014). Weighted distributions. *Wiley StatsRef: Statistics Reference Online* .

Preston, S.H. (1976). Family sizes of children and family sizes of women. *Demography* 13(1): 105–114.

Schoen, R. (2013). *Modeling Multigroup Populations.* Springer Science & Business Media.

Sen, P.K. (1987). What do the arithmetic, geometric and harmonic means tell us in length-biased sampling? *Statistics & probability letters* 5(2): 95–98.

Sheps, M.C. (1964). On the time required for conception. *Population Studies* 18(1): 85–97.

Sheps, M.C., Menken, J., and Radick, A.P. (1973). *Mathematical Models of Conception and Birth.* Chicago: University of Chicago Press.

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.

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

Appendix: Empirical example

In this appendix, we use the example of combining subnational life tables to show that appreciable errors can arise when death rates are not aggregated correctly. To do so, we create a pseudo-population based on the lifetables found in the US Mortality database.[2] The US Mortality database has life tables for each US State and for Washington, D.C. We treat these life tables as cohorts, each of which is initially the same size. We then compare three strategies for aggregating death rates for each single year of age, for each sex, and for both sexes combined.[3]

The aggregate death rate across these pseudo-cohorts is equal to

$$_nM_a = \frac{\sum_i {_nD_a^i}}{\sum_i {_nL_a^i}}. \tag{15}$$

where i indexes the pseudo-cohorts. In the main text, we show that this true aggregate is equal to $A_H[M, D]$ or, alternatively, $A_A[M, L]$.

In order to illustrate how much variation there is between the death rates across the subpopulations, we calculate the cofficient of variation:

$$\text{cv}_L\big(_nM_a^i\big) = \frac{\text{sd}_L\big(_nM_a^i\big)}{_nM_a}, \tag{16}$$

where $\text{sd}_L(_nM_a^i)$ is the standard deviation across the exposure-weighted state life table death rates. The coefficient of variation quantifies the amount of spread in death rates across the subnational units, when compared to the true aggregate death rate.

Figure 1 shows the aggregated age-specific death rates for the pseudo-population created from the 51 state lifetables (left panel); and the coefficient of variation across the 51 state lifetables (right panel).

For each age-sex group, we compare the true aggregate death rate to two alternate strategies for aggregating death rates. The first alternate strategy is the simple (unweighted) arithmetic mean of the state death rates:

---

[2]https://usa.mortality.org/

[3]For each age and sex, we restrict our sample to the states that have at least one death; this is typically all 51 states, but some small states have no deaths at low-mortality ages – Vermont, for example, has no deaths at age 2 in the 2015 USMD life table. The harmonic mean is undefined when the outcome includes values of zero, since the reciprocal of the outcome appears in the denominator of the harmonic mean.
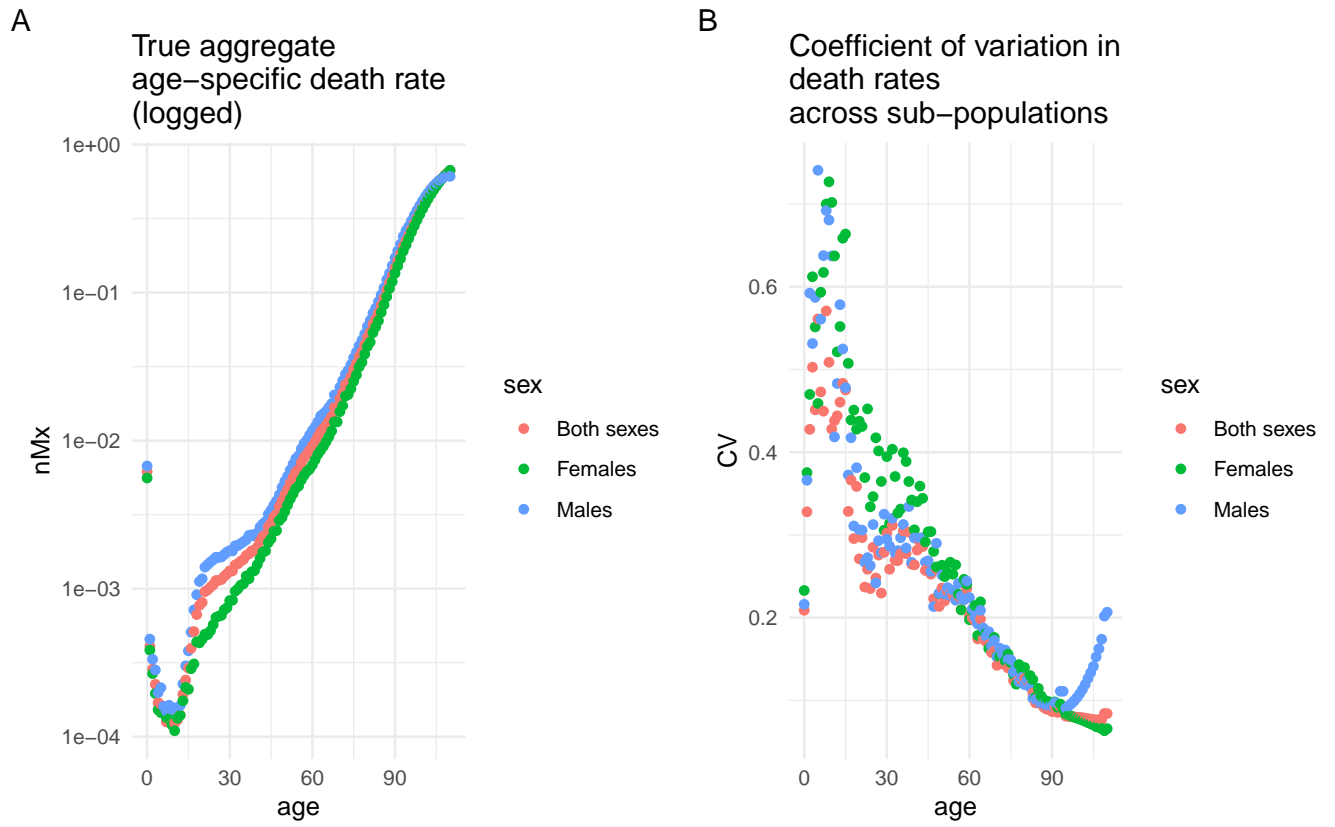
Figure 1: A) Aggregate death rates across sub-national psuedo-units; B) Relative variation in death rates across sub-national pseudo-units.
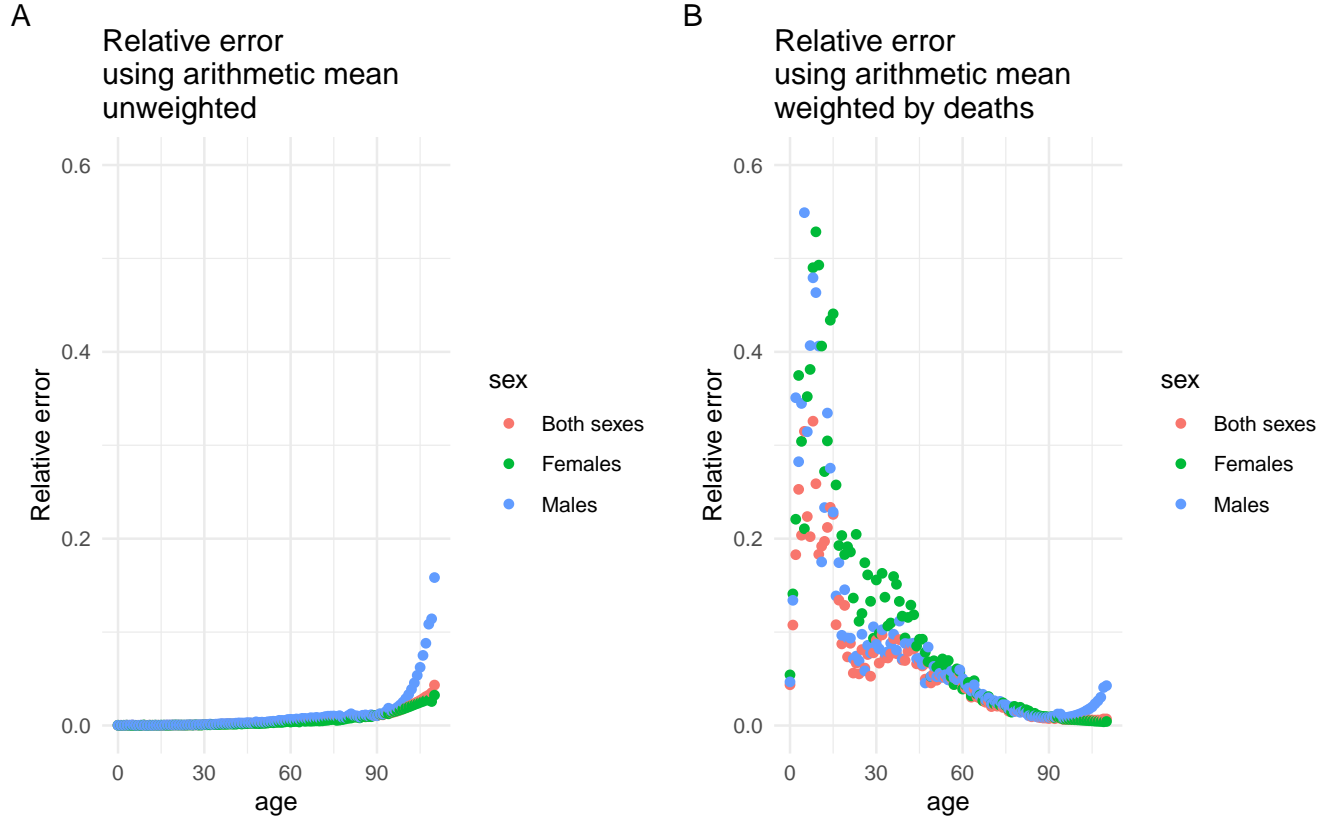
Figure 2: Relative errors in aggregating using the unweighted arithmetic mean (Panel A) or in using the incorrectly weighted arithmetic mean (Panel B).

$$_nM_a^I = \frac{1}{N} \sum_i {}_nM_a^i, \tag{17}$$

where N is the number of states.

The second alternate strategy is the arithmetic mean weighted by the number of deaths, rather than the exposure:

$$_nM_a^{II} = \frac{\sum_i {}_nM_a^i {}_nD_a^i}{\sum_i {}_nD_a^i}. \tag{18}$$

Figure 2 shows the relative errors, $\mathrm{re}(_nM_a^I) = \frac{_nM_a - _nM_a^I}{_nM_a}$ and $\mathrm{re}(_nM_a^{II}) = \frac{_nM_a - _nM_a^{II}}{_nM_a}$, that result from aggregating death rates under each strategy. The Figure reveals that relative errors can be of considerable magnitude – up to 10 percent or more in some scenarios.

For the death-weighted arithmetic mean, comparing panel B of Figure 2 to panel B of Figure 1

illustrates how relative errors in $_nM_a^{II}$ are greater when there is greater variation in the 51 underlying subpopulation death rates. To better understand this relationship between the relative errors in $_nM_a^{II}$ and the variation in the death rates, note that $_nM_a^{II}$, weights by deaths but incorrectly uses the arithmetic mean, instead of the harmonic mean. When weighting by the number of deaths, the arithmetic mean will be affected by length bias (Equation 9). To see this, note that the main text shows that the true aggregate death rate can be written as $A_A[M, L]$, or:

$$_nM_a = \frac{\sum_i {_nM_a^i}{_nL_a^i}}{\sum_i {_nL_a^i}} = \sum_i {_nM_a^i} \times \left(\frac{_nL_a^i}{_nL_a}\right),$$
(19)

where $_nL_a = \sum_i {_nL_a^i}$ is the aggregate exposure, and the term in parenthesis is the correct weight to use with the arithmetic mean. As we saw in the main text, taking the arithmetic mean of a length-biased sample would mean using weights that have the form $f^\star(y) = y \cdot f(y) \cdot (1/E_f[x])$. In this case, these length-biased weights are

$$_nM_a^i \cdot \frac{_nL_a^i}{_nL_a} \cdot \frac{1}{_nM_a} = \frac{_nM_a^i \, {_nL_a^i}}{_nL_a \, {_nM_a}} = \frac{_nD_a^i}{_nD_a}.$$
(20)

Thus, $_nM_a^{II}$, the death-weighted arithmetic mean of the subpopulation death rates, is a length-based average. In this case, Equation 8 shows that we should expect larger errors from $_nM_a^{II}$ when there is more variation in the death rates across the sub-populations. In fact, Equation 8 can be used to derive an exact expression for its relative error. In the main text, Equation 8 revealed that $E_{f^\star}[y] = E_f[x]\left[1 + \mathrm{cv}_f^2[x]\right]$. In our empirical setting, this relationship implies that $_nM_a^{II} = {_nM_a}\left[1 + \mathrm{cv}_L^2(_nM_a^i)\right]$. Thus, we have

$$\begin{aligned} \mathrm{re}(_nM_a^{II}) &= \frac{_nM_a^{II} - {_nM_a}}{_nM_a} \\ &= \frac{_nM_a\left[1 + \mathrm{cv}_L^2(_nM_a^i)\right] - {_nM_a}}{_nM_a} \\ &= \mathrm{cv}_L^2(_nM_a^i). \end{aligned}$$
(21)

Thus, the relative error in a length-biased aggregation of subpopulation death rates, $_nM_a^{II}$, is exactly equal to $\mathrm{cv}_L^2(_nM_a^i)$, the square of the coefficient of variation (Equation 16). Figure 3 shows this directly, confirming that in our empirical example the relative errors in $_nM_a^{II}$ (Figure 2, Panel B) are exactly equal to the squared coefficients of variation in the subpopulation death rates (Figure 1, Panel B).
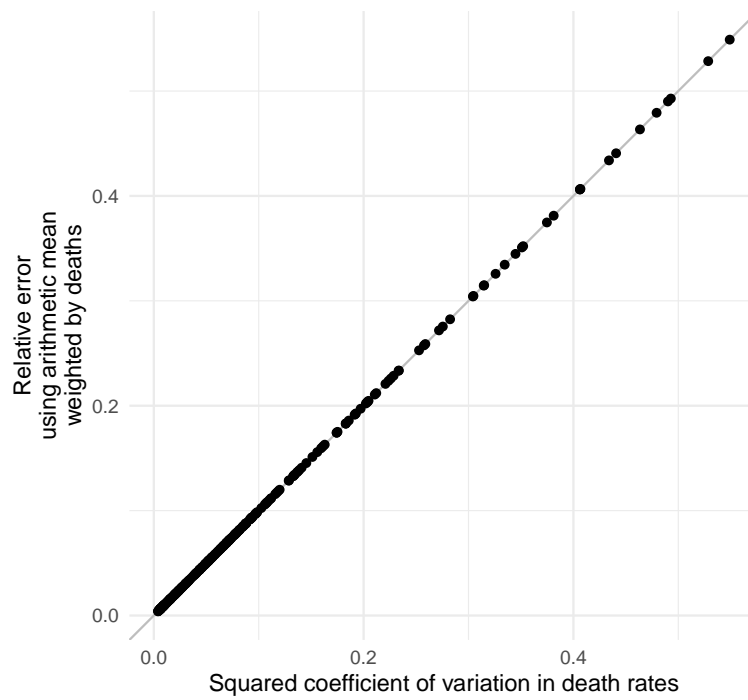
Figure 3: The relative errors resulting from incorrectly aggregating using the death-weighted arithmetic mean are exactly equal to the squared coefficient of variation in the subpopulation death rates. All points lie on the diagonal $y = x$ line, confirming that these two quantities are equal.