Chapter 3

Mortality of Supercentenarians: Estimates from the Updated IDL

Jutta Gampe

3.1 Introduction

Contributors to the IDL have made concerted efforts to collect data on individuals who survived to extreme ages that are free of age-ascertainment bias. The age of each of these individuals has been thoroughly validated. In their book on supercentenarians, Maier et al. (2010) described these efforts in detail, and the conclusions derived for human mortality beyond age 110 were presented in one of the chapters (Gampe 2010). The results were striking. Mortality was found to be flat after age 110, at a level corresponding to an annual probability of death of 50%. No differences between males and females could be detected, and no differences in levels of mortality were found between earlier and later birth cohorts.

Since then, the contributors to the IDL have continued their data collection and validation efforts. Consequently, the number of supercentenarians in the database has roughly doubled. The details of the IDL update and the specific data collection procedures are described in this volume (see Chap. 2). Here, the updated data are used to re-estimate the trajectory of human mortality beyond age 110.

3.2 Observation Schemes and Estimation Procedure

As survival to extreme ages is by definition rare, it is necessary to combine data from several countries to arrive at samples that are sufficiently large to allow us to make reliable inferences. The observation plans that are implemented depend on the available data in each country, and properly accounting for these sampling plans in the data analysis is crucial for generating unbiased results.

J. Gampe (✉)

Max Planck Institute for Demographic Research, Rostock, Germany

e-mail: gampe@demogr.mpg.de

© The Author(s) 2021

H. Maier et al. (eds.), *Exceptional Lifespans*, Demographic Research Monographs, https://doi.org/10.1007/978-3-030-49970-9_3

29

Two main observation plans can be found in the IDL. For countries with accurate population registers, both the number of supercentenarians who have died before a specific target date and the number who are alive at this date can be determined, which leads to observed ages at death as well as right-censored (still alive) observations. The other prominent observation scheme provides the deaths of all supercentenarians that occurred between two specific (period) dates. The consequences of such a sampling plan have been discussed in Gampe (2010), and the key figure is reproduced here for easier reference; see Fig. 3.1.

If only events (deaths) that occurred between two time points, t_1 and t_2 , are recorded, then individuals have to die ‘young enough’ – that is, before t_2 – to be included in the sample, while those who have been living ‘too long’ will not be seen in the data. As a consequence, we are not aware of individuals alive who are exposed to the risk of death during the sampling period, and this effect needs to be accounted for in the estimation procedure. This condition, whereby observations are included only if the event has occurred before the age when the individual would have left the sampling frame at time t_2 , is called right-truncation. Incorporating this sampling condition in the analysis is essential for obtaining unbiased estimates. As the number of supercentenarians has been growing in recent years, ignoring right-truncation can have substantial effects. Adjusting observations for general censoring and truncation patterns has a long tradition. This practice was discussed in Dempster et al. (1977); and, for the model we use here, in Pagano et al. (1994).

Variations of these two main sampling plans have been implemented in some countries, the details of which were discussed in Chap. 2 (Jdanov et al., this volume).

The second choice that has to be made is regarding the specific distributional model that is assumed for the variable of interest; here, life spans after age 110. While global parametric models are parsimonious and efficient, they determine the tail behavior or the distribution, and, hence, the hazard trajectory in the limit. By contrast, flexible (quasi-)nonparametric models in the spirit of life table analysis allow for the investigation of the hazard without such global assumptions. However, the flexibility in these models comes at the price of much greater variability in the estimates.

Fig. 3.1 Sampling frames, illustrating left- and right-truncation; from Gampe (2010)

The analysis strategy we pursue here is the same as the strategy used in Gampe (2010). The EM algorithm (Dempster et al. 1977) is employed to estimate a flexible model (Pagano et al. 1994) from the truncated and censored data. Standard errors for the estimates are not automatically provided by the EM algorithm, and several methods for estimating them have been suggested; see McLachlan and Krishnan (2008). Since the observed data information matrix (i.e., based on the incomplete data log-likelihood) can be derived analytically here without too much effort, this approach, which avoids numerical differentiation, was pursued. Standard errors for derived quantities, such as the survival function or the annual probabilities of death, are then determined by the delta method.¹

3.3 Results

In the previous analysis of the IDL data, the data from all countries were combined to achieve a sufficiently large sample size. The United States contributed a large share of the data in this analysis (341 out of 637 individuals). Obtaining separate estimates for different geographical regions is desirable. The increased number of individuals in the updated IDL allows us to conduct separate analyses for the U.S. on the one hand and the European countries on the other. We have also chosen to analyze the data for Japan separately, even though the sample size for this country is relatively small. Table 3.1 summarizes the three subsamples.

We first consider the sample for the United States. The estimates of the survival function for the flexible model, based on intervals of single years of age, is shown in the left panel of Fig. 3.2, with 95% confidence intervals added. The right panel shows the corresponding annual probabilities of death (again, with 95% confidence intervals). Uncertainty for the annual probabilities of death increases quickly, and is large after age 113, at which point the data are too sparse to allow for an accurate

Table 3.1 Summary of three regions analyzed separately (top) and the contributions of different countries to the European sample (bottom)

Country	USA					Europe					Japan	
Women	442					531					139	
Men	44					42					21	
Total	486					573					160	
Country	FRA	ITA	BEL	ESP	EW	DEN	SWE	NOR	FIN	AUT	CHE	GER
Women	160	125	20	51	121	4	11	8	5	6	4	16
Men	5	11	3	9	7	0	1	2	1	0	0	3
Total	165	136	23	60	128	4	12	10	6	6	4	19

¹Quadratic approximations to the log-likelihood can be problematic for estimates close to zero or one, leading to confidence intervals that extend beyond the unit interval for probabilities; alternative parametrization of the model via log-odds can mitigate this effect.

Fig. 3.2 Survival function (left) and the annual probability of death (right) for data from the U.S. Estimates for the flexible model are based on one-year intervals of age; the vertical bars give 95% confidence intervals. The red dashed line is the survival curve resulting from an exponential distribution, estimated by the maximum likelihood for the same data

Fig. 3.3 Survival function (left) and the annual probability of death (right) for the combined data from European countries. Estimates for the flexible model are based on one-year intervals of age; the vertical bars give 95% confidence intervals. The red dashed line is the survival curve resulting from an exponential distribution, estimated by the maximum likelihood for the same data

assessment. The flexible model is restricted to the range defined by the maximal observed age at death (which is 119 for the U.S.).

The shape of the survival curve suggests an exponential model, implying a constant hazard; and the resulting estimate is added in Fig. 3.2, with the parameter λ estimated by the maximum likelihood from the same data. The two survival curves are in (surprisingly) close agreement.

The estimated value $\hat{\lambda} = 0.7238$ (s.e. 0.0343) corresponds to an annual probability of death $\hat{q} = 0.5151$.

The same analysis was repeated with the IDL data from the European countries; the results are displayed in Fig. 3.3. Again, the level of agreement between the flexible model and the exponential model is striking, and the estimated hazard parameter is $\hat{\lambda} = 0.7953$ (s.e. 0.0365). The estimated parameter is slightly higher, but is

comparable to the estimate from the U.S. The corresponding estimated annual probability of death is $q^{\wedge} = 0.5486$.

The confidence intervals in Fig. 3.3 nicely illustrate the uncertainty implied by the right-truncated observations in the European data. The largest observation in the European data is for Jeanne Calment (at age 122), and the second largest observation is at age 116. Since there are no observed deaths in between these ages, the probability for these ages is practically zero. However, there is incomplete information on the possibly existing, but unobserved (because of right-truncation) supercentenarians who are still alive. This uncertainty about unobserved exposures is reflected in the stark increase in uncertainty for those ages.

Finally, the same analysis was performed for Japan. Although the sample is considerably smaller (160 individuals, 139 women, 21 men), Japan is of particular interest since it has been the record holder in life expectancy for many years. The results are presented in Fig. 3.4. The correspondence with the exponential distribution is less clear than it is for the U.S. and Europe, but if we estimate an exponential distribution, the estimated hazard is $\lambda^{\wedge} = 0.5891$ (s.e. 0.0481), which implies an annual probability of death $q^{\wedge} = 0.4452$.

Figure 3.5, left panel, summarizes the three estimates of $\lambda^{\wedge}$ for the U.S., Europe, and Japan. For Europe and the U.S., we estimated separate parameters for men and women; the resulting estimates of λ are given in Table 3.2. Although the point estimates are different, the uncertainty in the estimates (due to low numbers of male supercentenarians) does not allow us to reject the hypothesis that men and women have the same level of mortality; see also Fig. 3.5, right panel.

To investigate potential temporal trends in the level of mortality, both the European and the U.S. sample were split into early and late cohorts. The split was made so that the resulting subsets were of comparable size. Since the U.S. and the European data cover different birth cohort ranges, the split into early and late cohorts was done differently for the two samples. In the U.S. data, those individuals born

Fig. 3.4 Survival function (left) and the annual probability of death (right) for data from Japan. Estimates for the flexible model are based on one-year intervals of age; the vertical bars give 95% confidence intervals. The red dashed line is the survival curve resulting from an exponential distribution, estimated by the maximum likelihood for the same data

Fig. 3.5 Maximum likelihood estimates of the parameter λ assuming an exponential distribution for life spans after age 110. The horizontal bars give 95% confidence intervals. The details are summarized in Table 3.2

Table 3.2 Table of the maximum likelihood estimates of λ from different samples

Sample	MLE $\hat{\lambda}$	s.e. ($\hat{\lambda}$)
Europe	0.7953	0.0365
Europe Men	0.8617	0.1387
Europe Women	0.7898	0.0379
Europe Early	0.7629	0.0448
Europe Late	0.8520	0.0619
U.S.	0.7238	0.0343
U.S. Men	0.6537	0.1049
U.S. Women	0.7313	0.0362
U.S. Early	0.7609	0.0480
U.S. Late	0.6805	0.0495

before 1887 (252 observations) were included in the early cohort, while those individuals born in 1887 or later (234 observations) were included in the later cohort.

In the European data, those individuals born before 1896 (293 observations) were included in the early cohort, while those individuals born in 1896 or later (280 observations) were included in the later cohort.

The resulting parameter estimates are summarized in Table 3.2, and are displayed in Fig. 3.5, right panel. For both regions, there is no evidence that the mortality levels of the early and the later birth cohorts differ.

3.4 Discussion

The updated version of the IDL allows for a re-analysis of the life span distribution after age 110, and hence the trajectory of human mortality beyond this age. The proper incorporation of censoring and truncation patterns is essential for the

analysis. Estimates can be obtained via the EM algorithm for a flexible model that does not determine the global tail behavior of the distribution. The results of the model of constant mortality, implied by the exponential distribution, show strong agreement with the estimates obtained from the flexible model for both the European and the U.S. samples, in which the levels of mortality are found to be comparable.

Differences in mortality between the sexes cannot be detected in either sample due to the low numbers of male supercentenarians. A comparison of the mortality levels of earlier and later cohorts did not uncover significant differences in either the U.S. or the European dataset.

Thus, given these results, we confirm that the basic conclusions stated by Gampe (2010) still hold.

The picture for the data from Japan is less clear, as the mortality level for this dataset is lower, and the outcomes do not align as closely with the estimates of the exponential model. Because the sample size for Japan is much smaller, the conclusions are not as clear for this country as they are for Europe and the United States.

References

- Dempster, A. P., Laird, N. M., & Rubin, D. R. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.
- Gampe, J. (2010). Human mortality beyond age 110. In H. Maier, J. Gampe, B. Jeune, J.-M. Robine, & J. W. Vaupel (Eds.), *Supercentenarians* (pp. 219–230). Berlin: Springer.
- Maier, H., Gampe, J., Jeune, B., Robine, J.-M., & Vaupel, J. W. (Eds.). (2010). *Supercentenarians* (Demographic research monographs (A series of the Max Planck Institute for Demographic Research)). Berlin: Springer.
- McLachlan, G., & Krishnan, T. (2008). *The EM algorithm and extensions* (2nd ed.). Wiley.
- Pagano, M., Tu, X. M., Gruttola, V. D., & MaWhinney, S. (1994). Regression analysis of censored and truncated data: Estimating reporting- delay distributions and AIDS incidence from surveillance data. *Biometrics*, 50(4), 1203–1214.

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

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