LONGEVITY RECORDS ARE BOUNDLESS BUT RARE

ALESSANDRO FERRAIOLO

ABSTRACT

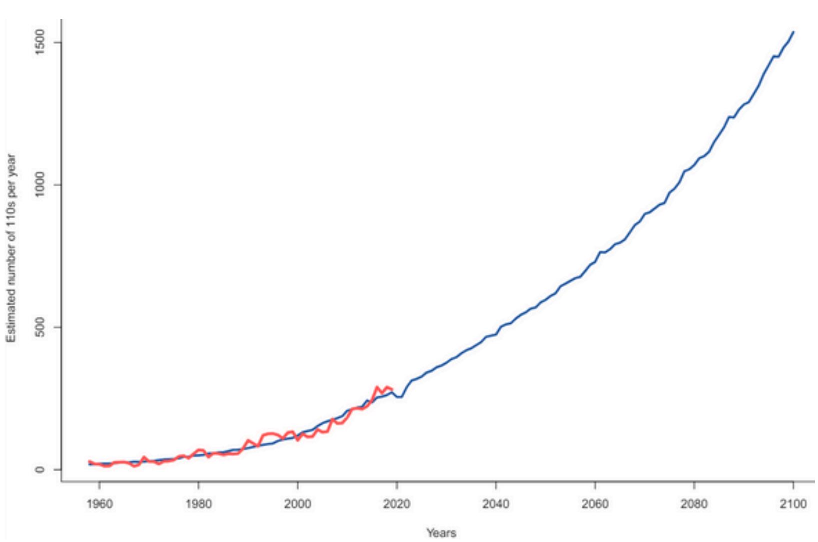
Our study deals with the estimation of new longevity records. It focuses on predicting the number of people aged 110, which is necessary to estimate when one of them will reach a record age. The attainment of an age greater than 110 is an exceedingly rare event. We estimate when we expect to find new records until 2116, based on the number of people aged 110. Our study operates under the assumption that yearly mortality rate beyond the age of 110 becomes constant at 0.5 [1]. Under this assumption, the probability of surviving an additional year after reaching age 110 is likened to a coin toss, where the outcome of each toss represents an extra year of life. Consequently, a higher number of tosses (representing individuals) increases the likelihood of at least one individual surpassing a certain age threshold. Given the number of supercentenarians in a certain period, we can assess the maximum age we expect at least one of them to reach. Given that data on individuals aged 110 and older are only available up to 2019, we initially developed a predictive model in excellent agreement with the experimental data to estimate the population of individuals aged 110 up to the year 2100. This estimation was derived using WHO's data on life expectancy and global population demographics. The data from about 110s up to 2019 was obtained from the HDM dataset.

THE DATASET - COUNTRY , YEAR , LIVING 110s

Since there is no available dataset meeting our needs of a constant, coherent, and complete number of living 110s over time, we estimated these data from the most reliable death count (however data was available only for 109s and not for 110s). To find the number of people who survive up to age 110, we have used the dataset "Death Counts" of HDM [2], which displays the number of people who died at age 109, year by year and country by country. Using the probability of dying between ages 109 and 110 (the mortality rate: H, included in the dataset), we estimated the number of people who survived up to age 110:

Number of living 110s (in a certain year and country) = number of dead 109s * (1-H) / H

From these findings, we constructed a robust dataset with three key columns: Country, Year, and Living110. The countries included were chosen for the consistency and reliability of their data over the period 1958-2019. We opted not to use the IDL dataset since the periods covered are different for each country [3]. Additionally, the IDL's stringent validation process for supercentenarians leads to an underestimation of the actual number. For this reason, IDL's dataset is unreliable for working with absolute quantities, and consequently, for estimating the future number of 110s.



[Figure 1]

On the graph the red line shows the observed data from 1958 up to 2019, while the blue line represents the predicted data from 1958 up to 2100. On the y-axis we report the estimated number of 110s per year.

THE MODEL TO PREDICT THE LIVING 110s

The yearly number of living 110s mainly depends on the total population and the life expectancy. Our predictive model for individuals aged 110 relies on the total population of the selected countries and their life expectancy. Initially, attempting a multilinear regression approach yielded unexpected results due to the simultaneous increase of both variables over time. To address this, we plotted the relative number of individuals aged 110 (#110s/Population) against life expectancy, ultimately selecting the probit function as the most suitable model. This decision was taken because the number of people dying at a certain age is almost normally distributed around the life expectancy. Therefore, as the life expectancy increases, the number of people dying at 109 first increases at an increasing rate, then at a decreasing rate. Because of their shape, the probit and logit models well describe this distribution. However, the second one is more appropriate for heavy-tailed distributions, which is not the case. Our analysis demonstrates the accuracy of the probit model through its alignment with observed data from 1958 to 2019.

- Observed 110s up to 2019 = 6155
- Predicted total 110s up to 2019 = 6153
- Margin of error = (6153-6155)/6153 = 0.0003

OUR FINDINGS

The conclusions drawn here presume that the count of supercentenarians restarted in 2020, hence the gap between previous and subsequent records might appear disproportionately large. If we had opted to initiate our predictions from 1958, the outcomes would have differed, but that would have been useless given that individuals aged 110 between 1958 and 2020 are now deceased. Jeanne Louise Calment (the only person with a documented lifespan of 122 years) represents an outlier, given that the expected number of people aged 122 until 1997 was 0.22 (based on available data). Therefore, that's a further justification to why we predict so much time will elapse before a new record.

References:
[1] Belzile, Davison, Gampe, Rootzén and Zholud (2022) Is There a Cap on Longevity? & Human life is unlimited – but short
[2] https://www.supercentenarians.org/en/data-and-metadata/
[3] https://www.mortality.org/Data/ZippedDataFiles

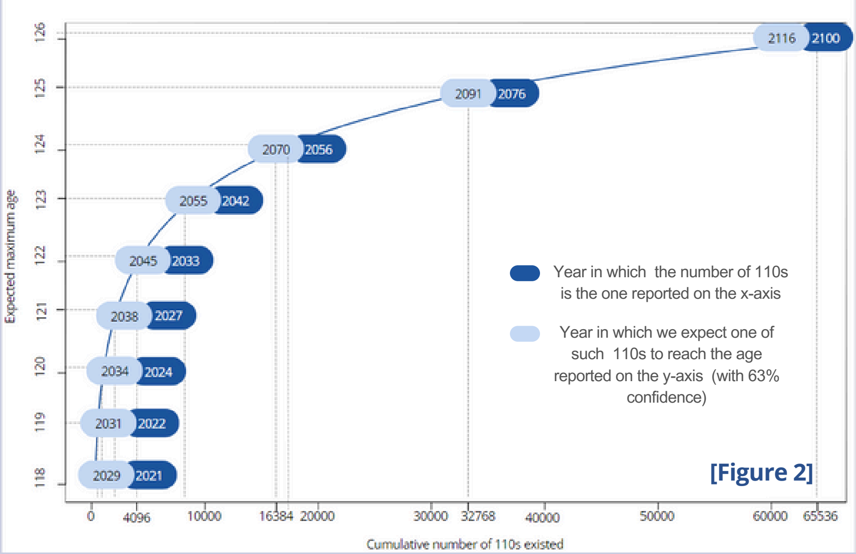
The expected number of people who reach age 110+n is given by the total number of people aged 110, times the probability to survive n years, which, given our assumptions, is (0.5)^n.
We consider the cumulative number of people eligible to beat the record starting from 2020 (which is the number of everyone who has reached the age of 110 since 2020):

$$\#[110+n] = \#110 * (0.5)^n$$

To determine the cumulative number of 110s necessary to expect one them to live for 110+n years (and beat a record), we set the number of people aged 110+n equal to 1 and solve for n:

$$n = log_{0.5}\left(\frac{1}{cum\,(\#110s)}\right)$$

Since n represents the years lived after 110, we add 110 back to get the actual age and then plot this function in **[Figure 2]**.



On the x-axis, we report the *cumulative* number of individuals we expect to reach age 110, starting from 2020. On the y-axis, we report n+110: the maximum age that one of those individuals is expected to reach.

[Figure 2] shows we expect one person to beat the current record and reach age 123 by 2055. Notice that reaching age 110+n is an exceedingly rare event (p = 0.5^n) . Therefore, when there is an adequate number of supercentenarians (the trials), the results can be described by a Poisson distribution. In particular, when the number of trials becomes 1/p, the expected amount of individuals reaching the desired age is 1, p*(1/p). Hence, the probability of the event to happen at least once is 1-dpois(0,1) = 0.6321. When the number of trials reaches ~ 3/p, the probability of the event to happen at least once is 1-dpois(0,3) ~ 0.95. Consequently, in **[Table 1]**, we calculate the year for which the probability for the event to have occurred at least once is 95%.

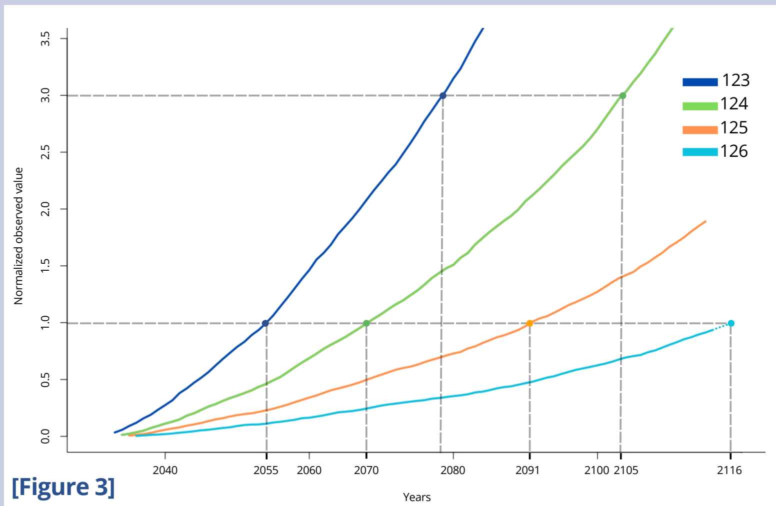
Age	p	n (63%)	year (63%)	n (95%)	year (95%)
122	0.5^12	4096	~2033 (+12)	~12260	~2049 (+12)
123	0.5^13	8192	~2042 (+13)	~24500	~2067 (+13)
124	0.5^14	16384	~2056 (+14)	~49000	~2091 (+14)
125	0.5^15	32768	~2076 (+15)	~96000	>2100 (+15)
126	0.5^15	65536	~2100 (+16)	~146000	>>2100 (+16)

[Table 1]

p = probability to reach "Age" provided a person has 110 years
n(63%) = requisite number of 110s to expect one person to reach "Age" with probability 63%.
year(63%) = year when the cumulative sum of people reaching age 110 becomes n(63%).
n(95%) = requisite number of 110s to expect one person to reach "Age" with probability 95%.
year(95%) = year when the cumulative sum of people reaching age 110 becomes n(95%).

Adding the number in parenthesis we get the year before which we expect to reach a new record with 63% (or 95%) likelihood.

To assess whether the predictions given by the model and the computations were consistent, we simulated multiple times the age reached by every individual aged 110 between 2020 and 2100. In doing so, we kept track of when the individuals reached age 110, and their year of death. We created a dataset, from which we could extract how many people reached a certain age in a determined year. Then for every age from 123 to 126 we looked at the cumulative sum year by year, checking when the cumulative sum became equal to the number of simulations. That year was mostly the exact same as year(63%) in **[Table 1]**.



The curves represent the cumulative sum of individuals reaching certain ages (123-126) throughout all simulations, divided by 1000 (the number of simulations). It is shown exactly when the curves reach value 1 (and 3 for 123 and 124), projecting the points on the x-axis to show the year. When a curve has value 1 we know the cumulative sum is equal to the number of simulations and therefore we expect one individual to have reached that age.