Supercentenarians and the oldest-old are concentrated into regions with no birth certificates and short lifespans

Saul Justin Newman*[1]

[1]Biological Data Science Institute, Australian National University

*Correspondence to: saul.newman@anu.edu.au

Short title: Supercentenarians are concentrated into regions with no birth certificates and short lifespans

Abstract

The observation of individuals attaining remarkable ages, and their concentration into geographic sub-regions or 'blue zones', has generated considerable scientific interest. Proposed drivers of remarkable longevity include high vegetable intake, strong social connections, and genetic markers. Here, we reveal new predictors of remarkable longevity and 'supercentenarian' status. In the United States, supercentenarian status is predicted by the absence of vital registration. The state-specific introduction of birth certificates is associated with a 69-82% fall in the number of supercentenarian records. In Italy, which has more uniform vital registration, remarkable longevity is instead predicted by low per capita incomes and a short life expectancy. Finally, the designated 'blue zones' of Sardinia, Okinawa, and Ikaria corresponded to regions with low incomes, low literacy, high crime rate and short life expectancy relative to their national average. As such, relative poverty and short lifespan constitute unexpected predictors of centenarian and supercentenarian status, and support a primary role of fraud and error in generating remarkable human age records.

Introduction

The concentration of remarkable-aged individuals within geographic regions or 'blue zones' [1] has stimulated diverse efforts to understand factors driving survival patterns in these populations [2,3]. Both the overall population residing within these regions, and the individuals exceeding remarkable age cut-offs, have been subject to extensive analysis of lifestyle patterns [2,4–6], social connections [3,7], biomarkers [8,9] and genomic variants [10], under the assumption that these represent potential drivers behind the attainment of remarkable age.

However, alternative explanations for the distribution of remarkable age records appear to have been overlooked. Previous work has noted the potential of population illiteracy [11] or heterogeneity [12] to explain remarkable age patterns. More recent investigations revealed a potential role of errors [13–16] and operator biases [17] in generating old-age survival patterns and data. In turn, these findings prompted a response with potentially disruptive implications: that, under such models, the majority if not all remarkable age records may be errors [18].

Here, we explore this possibility by linking civil registration rates to per-capita estimates of remarkable age attainment, obtained from central population registries and validated supercentenarian databases in the USA and Italy.

These data reveal that remarkable age attainment is predicted by indicators of error and fraud, including illiteracy, poverty, high crime rates, short average lifespans, and the absence of birth certificates. As a result, these findings raise serious questions about the validity of an extensive body of research based on the remarkable reported ages of populations and individuals.

Methods

The number and birthplace of supercentenarians, individuals attaining 110 years of age, were downloaded from the Gerontology Research Group supercentenarian tables (updated 2017) and split into subnational units for birth locations (S1 Code). Populations were excluded due to incomplete subnational birthplace records (<25% complete) or poor subnational resolution (<15 total provinces), leaving only US supercentenarian population data.

Supercentenarians from the USA were matched to the 1900 survey counts for state and territory populations [19], and linked to the National Center for Health Statistics estimates for the timing of complete birth and death certificate coverage in each US state and territory [20]. Both the number of supercentenarian births overall, and estimates of supercentenarians per capita, approximated by dividing supercentenarian number by state population size in the 1900 US census [19], were averaged across the USA and represented as discontinuity time series relative to the onset of complete-area birth registration (S1 Code).

To quantify the distribution of remarkable-aged individuals in Italy, province-specific quinquennial life tables were downloaded from the Italian Istituto Nazionale di Statistica Elders.Stat database [21] to obtain age-specific survivorship data (S1 Code). Using cross-sectional data across Italian provinces, rates of age-specific survivorship (l_x) for ages 90-115 and life expectancy at age 100 were fit as dependent variables, and survival rates at age 55 and life expectancy from age 55 onwards as independent variables, using simple linear regression (S1 Code).

Results

The introduction of complete vital registration in the USA coincided with this rapidly increase in lifespan and population size, and was expected to result in a large increase in the number of supercentenarian records per capita. From 1880 to 1900, the core survey period for supercentenarians, the US population increased by 150% and average life expectancy around 20% [19,22].

Instead, the introduction of state-wide birth certification coincides with a sharp reduction in the number of supercentenarians. In total, 82% of supercentenarian records from the USA (N=536) predate state-wide birth certification. Forty-two states achieved complete birth certificate coverage during the survey period. When these states transition to state-wide birth registration, the number of supercentenarians falls by 80% per year (Fig 1a), or approximately 69% per capita (Fig 1b).

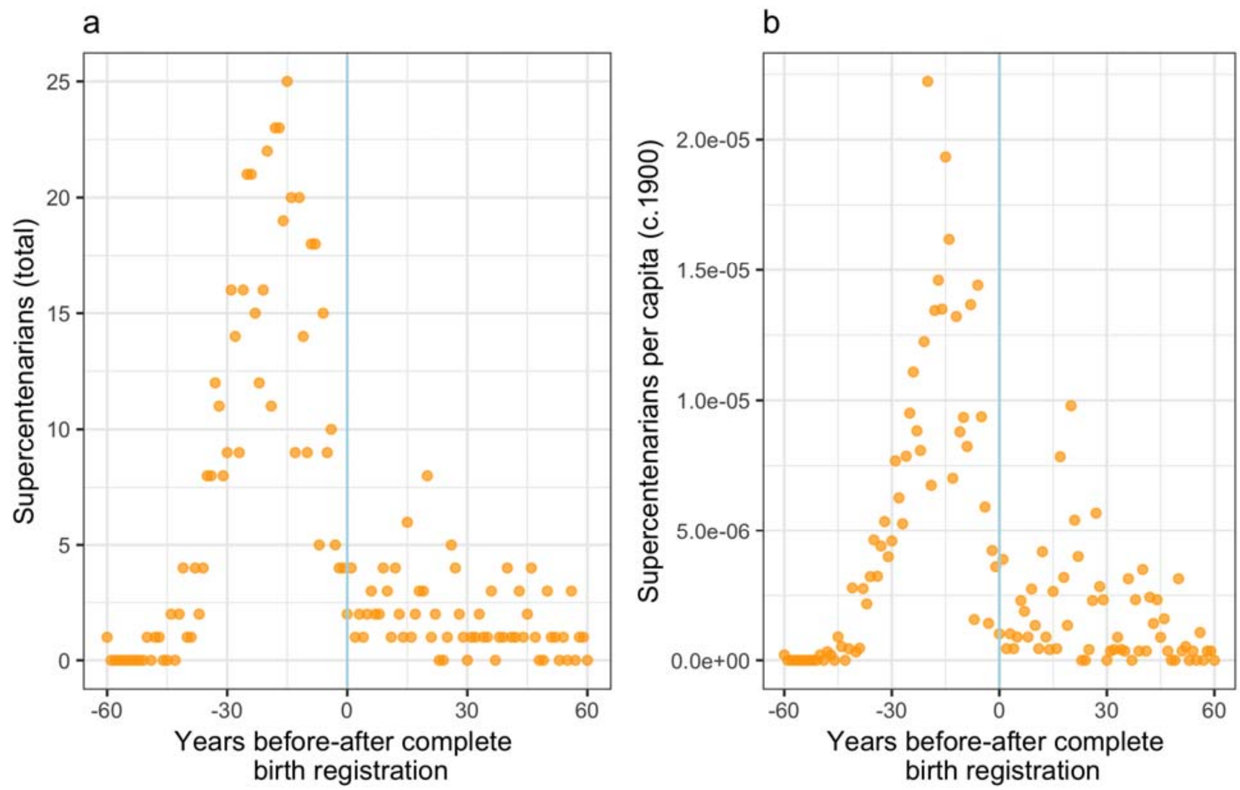

Figure 1. Number and per capita rate of attained supercentenarian status across US states, relative to the introduction of complete-area birth registration. Despite the combined effects of rapid population growth and increasing life expectancy during this period (c.1814-1904), the total number of US supercentenarians (a) falls dramatically after the introduction of state-wide birth certificates (vertical blue line). This trend remains after adjusting for total population size c.1900 (b) within each state.

The introduction of birth certificates in Italy largely predates the onset of supercentenarian records. Instead, the attainment of remarkable age in Italy is predicted by a short average lifespan. In Italy higher early- and mid-life survival is inversely correlated with mortality rates after age 95 (Fig 2a). Cohort survival to age 55 is negatively correlated with survival to ages 100 (Fig. 2b), 105 (Fig 2c) and 110 years (Fig 2d), and with life expectancy at age 100 (r = -0.4; p=0.00001).

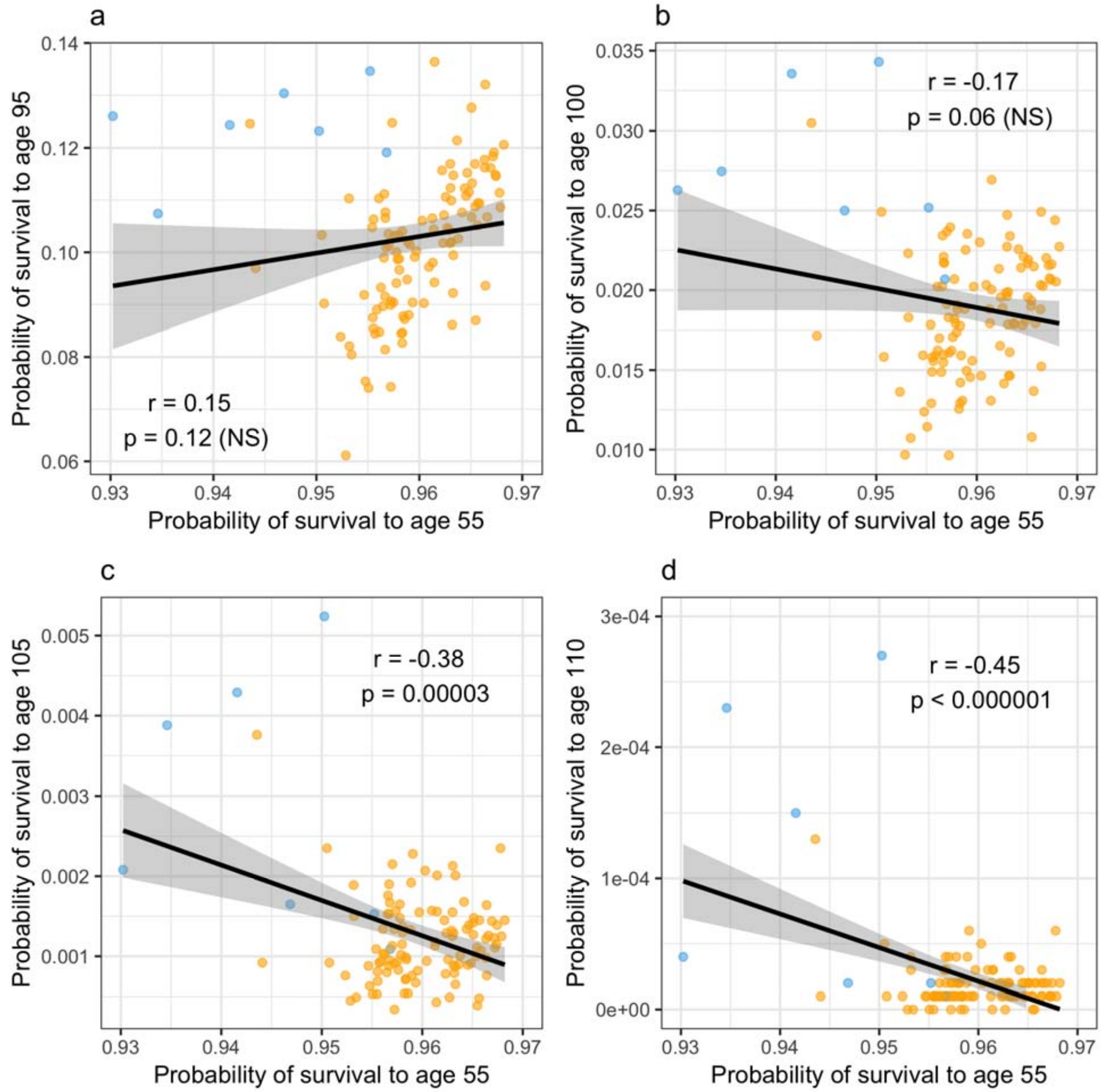

Figure 2. Relationship between mid-life and late-life survival across Italian provinces. Rates of survival during mid-life are positively correlated with survival at older ages across Italian provinces (points) until around age 95 (a; r = 0.15; p=0.1; N=116). However, this relationship inverts at advances ages: better mid-life and early-life survival rates, and higher average longevity, are linked to significantly smaller odds of surviving past 100 years (b), 105 years (c), or age 110 years (d). Sardinian provinces shown in blue.

These Italian and US data have previously been presented as hallmark ‘clean’ datasets of the oldest-old. However, there are strong indications that vital registration errors and predictors of error play substantial roles in the aggregation of remarkable age records. The US data support the hypothesis that improved vital registration should reduce the number of supercentenarians, and be associated with changing patterns of old-age survival, by reducing age-coding error rates [16]. Likewise, findings from the Italian data support the hypothesis that these ‘semisupercentenarians’ largely constitute a collection of age reporting errors [18].

Discussion

Italians over the age of 100 are concentrated into the poorest, most remote and shortest-lived provinces, while US supercentenarians are concentrated into populations with incomplete vital registries. Both patterns are difficult to explain through biology, but are readily explained as economic drivers of pension fraud and reporting error.

These results may reflect a neglect of error processes as a potential generative factor in remarkable age records. For example, Okinawa has the highest number of centenarians per capita of any Japanese prefecture and remains world-famous for remarkable longevity. Okinawa also has the highest murder rate per capita, the worst over-65 dependency ratio, the second-lowest median income, and the lowest median lifespan of all 47 Japanese prefectures [23].

Like the 'blue zone' islands of Sardinia and Ikaria, Okinawa also represents the shortest-lived and second-poorest region of a rich high-welfare state. These regions may have higher social connections and vegetable intakes, yet they also rank amongst the least educated and poorest regions of their respective countries. The hypothesis that these relatively low literacy rates and incomes are generating age-reporting errors and pension fraud, and therefore remarkable age records, seems overlooked.

Indicators of poverty and fraud, and contra-indications of health, are also ignored in remarkable-age surveys. For example, smoking rates of *e.g.* 17-50% [6] and illiteracy rates of 50-80% [5,6] are often observed in samples of the oldest-old. Surveying the 'blue zone' of Ikaria, Chrysohoou *et al.* observed that the oldest-old have: a below-median wage in over 95-98% of cases, moderate to high alcohol consumption (5.1-8.0 L/ year), a 10% illiteracy rate, an average 7.4 years of education, and a 99% rate of smoking in men [3].

Instead of prompting skepticism, under the relatively safe assumption that smoking, drinking, poverty, and illiteracy should not enrich for remarkable longevity records, these contra-indications of survival are routinely ignored. In contrast, it could be suggested that the abundance of supercentenarians in these regions reflect high rates of undetected error.

High-quality universal registration systems often contain undetected high-frequency errors. For example, contrary to previous assertions that "*Japan has…among the highest quality data for the oldest-old*" [24], a 2010 investigation of Japanese records revealed that 238,000 centenarians were actually missing or dead [25].

Data cleaning and error correction are the main approach to combat age-coding errors and enriching samples for real cases. Validation of the Japanese centenarian data result in a shift in population patterns, eliminating a clear link between centenarian abundance, low income and low mean longevity. Likewise, these relationships are absent from 'cleaned' centenarian data in the USA and the UK [26].

However, while clearly useful, these methods often produce the mistaken impression that the resulting 'validated' data are therefore largely free from error. For example, the Italian and US data used in this study show patterns consistent with a high frequency of type I age-coding errors. However, both populations were already subjected to extensive cleaning and validation [27], and are widely considered high-quality data and assumed to be 'clean'.

The logic behind these assumptions is informative. Post-validation errors in these Italian data were assumed to be minimal on the basis of a belief [27]. Subsequently, it was acknowledged that an unknown number of errors in these data could not be detected using documentary evidence, as "*Occasionally…a mistake will escape even a rigorous validation procedure*" [18]. Finally, it was proposed that the occurrence of such errors, which cannot be detected using documents, must be rare or "*essentially impossible*", because of the high quality of documents used to compile these data [18].

This argument might have been countered by another opinion: that a handwritten century-old database containing millions of entries, no independent biological validation, and an unknown type I error rate, might easily generate the few hundred annual errors required for a supercentenarian database. However, this criticism would ignore a more fundamental problem.

Physical possession of valid documents is not an age guarantee. Consider a room containing 100 real Italian supercentenarians, each holding complete and validated documents of their age. One random centenarian is then exchanged for a younger sibling, who is handed their real and validated birth documents. How could an independent observer discriminate this type I substitution from the 99 other real cases, using only documents as evidence?

This hypothetical error cannot be excluded on the basis of document consistency: every document in the room is both real and validated. In addition, a real younger sibling is also likely to have sufficient biographic knowledge to pass an interview. As such, any similar substitution error has the potential to indefinitely escape detection.

This 'Italian sibling' thought experiment reveals why type I age-coding errors cannot be ruled out, or even necessarily measured, on the basis of documentary evidence. It also reveals how debates on the frequency of these errors are not driven by direct empirical measurements, but by inference and opinion.

This issue presents a substantial problem for remarkable-age databases, embodied in a deliberately provocative, if seemingly absurd, hypothesis:

Every 'supercentenarian' is an accidental or intentional identity thief, who owns real and validated 110+ year-old documents, and is passably good at their job.

This hypothesis cannot be invalidated by the further scrutiny of documents, or by models calibrated using document-informed ages [28,29]. Rather, invalidating this hypothesis requires a fundamental shift: it requires the measurement of biological ages from fundamental physical properties, such as amino acid chirality or isotopic decay [30].

Until such document-independent validation of remarkable ages occurs, the type I error rate of remarkable human age samples will remain unknown, and the validity of ‘supercentenarian’ data in question.

Acknowledgements

The author would like to acknowledge Zoe Campbell for providing much-needed feedback.

References

1. Buettner D, Skemp S. Blue Zones: Lessons From the World's Longest Lived. American Journal of Lifestyle Medicine. 2016. doi:10.1177/1559827616637066
2. Poulain M, Pes GM, Grasland C, Carru C, Ferrucci L, Baggio G, et al. Identification of a geographic area characterized by extreme longevity in the Sardinia island: The AKEA study. Exp Gerontol. 2004; doi:10.1016/j.exger.2004.06.016
3. Chrysohoou C, Panagiotakos DB, Siasos G, Zisimos K, Skoumas J, Pitsavos C, et al. Sociodemographic and lifestyle statistics of oldest old people (>80 years) living in ikaria island: The ikaria study. Cardiol Res Pract. 2011; doi:10.4061/2011/679187
4. Rajpathak SN, Liu Y, Ben-David O, Reddy S, Atzmon G, Crandall J, et al. Lifestyle factors of people with exceptional longevity. Journal of the American Geriatrics Society. 2011. doi:10.1111/j.1532-5415.2011.03498.x
5. Afonso RM, Ribeiro O, Vaz Patto M, Loureiro M, Loureiro MJ, Castelo-Branco M, et al. Reaching 100 in the Countryside: Health Profile and Living Circumstances of Portuguese Centenarians from the Beira Interior Region. Curr Gerontol Geriatr Res. 2018;2018. doi:10.1155/2018/8450468
6. Kwon IS, Kim C-H, Ko HS, Cho S Il, Choi YH, Park SC. Risk factors of cardiovascular disease in Korean exceptional longevity. J Korean Geriatr Soc. 2005;9: 251–265.

7. Franceschi C, Bonafè M. Centenarians as a model for healthy aging. Biochem Soc Trans. 2004; doi:10.1042/bst0310457

8. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci. 2012; doi:10.1073/pnas.1120658109

9. Mondello C, Petropoulou C, Monti D, Gonos ES, Franceschi C, Nuzzo F. Telomere length in fibroblasts and blood cells from healthy centenarians. Exp Cell Res. 1999;248: 234–242. doi:10.1006/excr.1999.4398

10. Sebastiani P, Solovieff N, Puca A, Hartley SW, Melista E, Andersen S, et al. Genetic Signatures of Exceptional Longevity in Humans. Science (80-). 2010; doi:10.1126/science.1190532

11. Fries JF. Aging, natural death, and the compression of morbidity. Bulletin of the World Health Organization. 2002. pp. 245–250. doi:10.1056/NEJM198007173030304

12. Vaupel JW, Manton KG, Stallard E. The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. Demography. 1979;16: 439. doi:10.2307/2061224

13. Gavrilov LA, Gavrilova NS. Late-life mortality is underestimated because of data errors. PLoS Biol. 2019;17: e3000148. doi:https://doi.org/10.1371/journal.pbio.3000148

14. Preston SH, Elo IT, Stewart Q. Effects of age misreporting on mortality estimates at older ages. Popul Stud (NY). 1999; doi:10.1080/00324720308075

15. Gavrilova NS, Gavrilov LA. Mortality Trajectories at Extreme Old Ages: A Comparative Study of Different Data Sources on U.S. Old-Age Mortality. Living to 100 Monogr. 2014;2014. doi:10.1126/scisignal.2001449.Engineering

16. Newman SJ. Errors as a primary cause of late-life mortality deceleration and plateaus.

PLoS Biol. 2018;16: e2006776. doi:https://doi.org/10.1371/journal.pbio.2006776

17. Newman SJ. Plane inclinations: A critique of hypothesis and model choice in Barbi et al. PLoS Biol. 2018;16: e3000048. doi:https://doi.org/10.1371/journal.pbio.3000048
18. Wachter KW. Hypothetical errors and plateaus: A response to Newman. PLoS Biol. 2018;16: e3000076. doi:https://doi.org/10.1371/journal.pbio.3000076
19. Farnsworth Riche M, Benton B, Schnelder PJ, Norton AJ. Population of the States and Counties of the United States: 1790 to 1990 [Internet]. Washington D.C.; 1996. Available: https://www.census.gov/population/www/censusdata/PopulationofStatesandCountiesoftheUnitedStates1790-1990.pdf
20. Hetzel AM. History and Organization of the Vital Statistics System. US Vital Statistics System: Major Activities and Developments, 1950-95. Centers for Disease Control and Prevention; National Center for Health Statistics; 1997. p. 75. Available: https://www.cdc.gov/nchs/data/misc/usvss.pdf
21. Elders.Stat database: Istituto Nazionale di Statistica. In: Istituto Nazionale di Statistica [Internet]. [cited 7 Feb 2019]. Available: http://dati-anziani.istat.it/
22. Gibson C, Jung K. Historical Census Statistics On Population Totals By Race, 1790 to 1990, and By Hispanic Origin, 1970 to 1990, For Large Cities And Other Urban Places In The United States. Washington D.C.; 2005.
23. Director-General for Statistics and Information Policy, Ministry of Health L and W (organisation). Vital Statistics [Internet]. 2019 [cited 7 Feb 2019]. Available: https://www.mhlw.go.jp/english/database/db-hw/index.html
24. Willcox DC, Willcox BJ, He Q, Wang NC, Suzuki M. They really are that old: A validation study of centenarian prevalence in Okinawa. Journals Gerontol - Ser A Biol Sci

Med Sci. 2008; doi:10.1093/gerona/63.4.338

25. Japanese Ministry of Justice. About family register office work to affect location unknown elderly people [Internet]. 2010 [cited 18 Jun 2019]. Available: http://www.moj.go.jp/MINJI/minji04_00008.html.
26. Evans CJ, Ho Y, Daveson BA, Hall S, Higginson IJ, Gao W. Place and Cause of Death in Centenarians: A Population-Based Observational Study in England, 2001 to 2010. PLoS Med. 2014; doi:10.1371/journal.pmed.1001653
27. Barbi E, Lagona F, Marsili M, Vaupel JW, Wachter KW. The plateau of human mortality: Demography of longevity pioneers. Science (80-). 2018;360: 1459–1461.
28. Fleischer JG, Schulte R, Tsai HH, Tyagi S, Ibarra A, Shokhirev MN, et al. Predicting age from the transcriptome of human dermal fibroblasts. Genome Biol. 2018; doi:10.1186/s13059-018-1599-6
29. Malli RC, Aygun M, Ekenel HK. Apparent Age Estimation Using Ensemble of Deep Learning Models. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2016. doi:10.1109/CVPRW.2016.94
30. Nielsen J, Hedeholm RB, Heinemeier J, Bushnell PG, Christiansen JS, Olsen J, et al. Eye lens radiocarbon reveals centuries of longevity in the Greenland shark (Somniosus microcephalus). Science (80). 2016;353: 702–704. doi:10.1126/science.aaf1703