

Rejoinder to discussion of the paper “Human life is unlimited – but short”

Holger Rootzén and Dmitrii Zholud*

Abstract

What can be learned from data about human survival at extreme age? In this rejoinder we give our views on some of the issues raised in the discussion of our paper Rootzén and Zholud (2017).

Keywords: Extreme human life lengths, supercentenarians, IDL, GRG, Jeanne Calment, truncation, censoring, limit for human lifespan, age-biased sampling, generalized Pareto distribution, LATool.

1 Introduction

We thank the discussants¹ for very stimulating, thought-provoking, and educational comments. We were impressed by the title of Davison’s contribution, and by the precise prediction in the bible cited by Stoev & Battacharya, and by their attempt to “play God’s advocate”.

2 Biology, accident deaths, compression of morbidity

Nerman gives a quick and useful pointer to the very large literature on biological theories of aging, and writes that the question of a limit for human lifespans is not primarily statistical, but biological. We agree with this, but also think that one should use available data as efficiently as possible to give an empirical underpinning of biological theories – and also because of the intrinsic interest of the problem.

In their impressive contribution – a full paper on its own – Stoev & Battacharya make the intriguing comment that not knowing the cause of death may lead to bias. E.g., if the roof of the home of a supercentenarian falls down and kills all of the inhabitants, the observed supercentenarian life length should perhaps be considered as censored rather than fully observed. In a less dramatic, and often occurring event, if a supercentenarian has a fall which shortens her life length, should this be taken into account in the analysis? It could perhaps have been avoided by changing the layout of the home. But on the other hand the fall is usually also an effect of the frailty of the supercentenarian. Should the answer to the question about a biological limit for the human lifespan aim at describing life lengths of humans living in a “test tube” where no falls are possible and where they will not die from infectious diseases? But, could any human live like this? In the end, perhaps the most interesting approach is

**Department of Mathematical Statistics*

Chalmers University of Technology and University of Göteborg, Sweden.

E-mails: hrootzen@chalmers.se and dmitrii@chalmers.se

¹Davison (2018), Ferreira and Huang (2018), Keiding (2018), Nerman (2018), Segers (2018), Stoev and Battacharya (2018), and Zhou (2018). John Einmahl, Jutta Gampe, and Jan Vijn were also invited to participate in the discussion.

still to study lifespans as they are observed under the biological and cultural circumstances which the supercentenarians have lived under. However, if one takes cause of death into account in a statistical analysis this would amount to changing some observations from truncated to truncated and censored, and would lead to a longer estimated lifespan.

We agree with the hope expressed by Stoev & Battacharya that extreme value statistics could contribute to the quite difficult statistical analysis surrounding the question whether we will “age healthy” or “age sick”. This discussion was started by the “compression of morbidity” hypothesis of Fries (1980). His arguments build on an assumed finite limit for the human lifespan, but an unbounded lifespan is also compatible with both compression and expansion of morbidity. Fries inferred a finite limit of life lengths from an ideal “rectangularization” of survival curves, from a projected upper limit of 85 years for life expectancy, to be achieved in 2045, and from the fact that at the time when his paper was written the largest known human lifespan was 114 years. However, the ideal rectangular shape of the survival curve in Fig. 1 in his paper is contradicted by the very dramatic increase in survival up to age 100; see Vaupel (2010). Further, in 2015 the expected life length for women in Japan was 86.8 year, well above the Fries limit; finally the IDL data base contains 10 humans validated at level A and living longer than 115 years and the version of the GRG data base used in our paper contains an additional 14, with several added later.

3 Practical extreme value statistics

To address comments by Nerman, Segers, Stoev & Batacharya, and Zhou: The goals of an extreme value statistics analysis more often than not are both to increase the understanding of the extreme events, say, extreme life lengths, and to extrapolate the distribution of event sizes a bit outside the range of observations – but never to extrapolate all the way to infinity. Our practical approach to this is to find the simplest possible model for the extreme observations at hand, and then use it for understanding and extrapolation.

Occam’s razor is the classical expression of “simple”, and “simple” is also expressed in Einstein’s adage “Raffiniert ist der Herrgott, aber boshaft ist er nicht”. This is the hope that in the absence of information to the contrary simple models are those which describe our world most usefully. Simple models increase understanding: One learns from the ways data agree with or deviate from the simple model, and learning increases if different researchers start by trying the same simple model, as opposed to when everyone uses a different complicated model.

“Simple” means different things in different contexts. For excesses of high thresholds, the simplest model is that they follow an exponential distribution: then excesses of even higher threshold have the same exponential distribution, and the exponential distribution (of course, and for many parent distributions) occurs as the limiting distribution of scale normalized excesses. Assuming an exponential distribution is the default in statistical reliability theory. The second simplest one is the family of generalized Pareto (GP) distributions. For GP distributed excesses, excesses of a higher threshold also follow a scale changed version of the same GP distribution, and the GP distributions is the family of distributions which can be obtained as limits of distributions of threshold excesses. These characterizations are completely parallel to the properties of the normal distribution which make it the simplest distribution in non-extreme statistics. The next level of generality could then be to include covariates in the parameters, or to use second order regular variation to construct a more general family of distributions, or ...

For the IDL supercentenarian data the simplest model is an exponential distribution for excess ages, without any influence of the covariates sex, time, or group of countries, as checked by embedding it in the family of generalized Pareto distributions and by testing for non-exponentiality and for inclusion of covariates. A further crucial confirmation of the simple model is given by the nonparametric analysis in Gampe (2010). The model that survival after age 110 is exponential thus constitutes what we can

learn from existing data, and what can be used for extrapolation. As always extrapolation beyond the range of data comes with caveats, as discussed below.

Nerman does not see any convincing reason to restrict analysis of excess life length data by assuming that they follow a generalized Pareto distribution, and is not convinced by the extrapolation to the age range 120-130 years. Above we have set out our reasons for disagreeing with Nerman’s first point. But we think Nerman is right in his comment about extrapolation: Data convincingly shows an exponential distribution of survival for ages 110 – 115 and indicate that for ages 116 – 122 survival is also exponential. For ages 123 – 130 there is no data, and reality could turn out to be different. Extrapolation to these ages is still useful, we believe, because of its intrinsic interest, and because it makes it possible to detect interesting changes in survival as quickly as possible. And then, to extrapolate one should use the simplest model.

In contrast to Nerman, Stoev & Battacharya write “Extreme Value Theory is the most natural framework that can provide a principled answer to the question about whether or not natural human lifespan is finite”. Davison, Segers, and Zhou also use this framework, but raise a number of questions related to our analysis.

Under the heading “Uncertainty quantification”, Segers assumes that data follows a generalized Pareto distribution and discusses the issue that from finite data one can never be sure that a parameter of this distribution has a specific value, say 0. A general version of this problem is that if a smaller statistical model is continuously embedded into a larger one, then from observing a finite number of values one can never be sure that the smaller model is the right one. An extreme and unwanted conclusion from this argument would be that model selection, one of the most important tools of applied statistics, is invalid and that one always should use the largest model one could imagine. For some discussion of this issue, see Section 3.2 of our paper. We found the philosophical arguments in Mayo and Cox (2006) helpful.

Stoev & Battacharya address the same issue as Segers from a different angle by using “testing affinity” to quantify the statistical difficulty of the question of finiteness or not of the human lifespan. Their conclusion is that the amount of data so far available may not be sufficient to give a very confident answer to the question. Similarly, Zhou, using expected information rather than observed information and assuming non-truncated observation, notes that for the number of observations in the IDL database, an estimate of -0.082 or lower of the shape parameter γ has to be obtained before the null hypothesis $\gamma = 0$ can be rejected. A similar way to treat the same issue, briefly mentioned in our paper, is through power calculations.

Zhou makes the most detailed use of extreme value theory by assuming that human life lengths belong to the domain of attraction of an extreme value distribution with a second order index ρ , and writes that then the optimal sample fraction, k , to use is $O(n^{2\rho/(2\rho-1)})$, where n is the total number of observations. The total number of deaths, n , in the countries and time periods included in the IDL data is of the order 10^8 (so very likely the IDL data is the most extreme one any of us has seen). Solving the equation $n^{2\rho/(2\rho-1)} = 566$ one can see that as soon as the second order index is less than $-.26$ then the IDL sample size is smaller than what would be optimal, and hence that bias does not dominate. However, to use calculations like this one for practical statistics is carrying mathematics too far, we believe. Instead second order variation could be seen as a way to construct more general models that include the generalized Pareto models.

Further, from our practical point of view, Zhou’s comment about the existence of distributions which have a finite endpoint but with asymptotically exponential threshold excesses is irrelevant. We do not try to find a γ which lives all the way out in asymptotia, but use asymptotic reasoning to suggest suitable models for the data which has been observed.

In conclusion, the comments about the limited statistical resolution of the IDL – or any – data set are relevant and have to be kept in mind when using our results (as also discussed in our paper). Similarly, one never knows if a prediction outside of the range of the data will hit the mark. But

available data does not give any reason to come to any other conclusion than that survival after age 110 is exponential, so that human life is unlimited but short.

4 Confidence intervals and GP fitting to data covering lower ages

Davison used the IDL validation level A data (using also the parts of the US and Japan data which were excluded in our paper) to provide profile likelihood confidence intervals for the endpoints of the fitted GPD distributions. He obtained intervals which all contain infinity and with relatively high lower limits, and made the remark that these intervals probably are conservative.

Stoev & Battacharya used new statistical technology developed in their contribution to provide confidence intervals for the endpoint of the distribution of lifespans, and only used the 100 or 200 longest lifespans. Their intervals are built on regular variation at a finite endpoint, and are similar, but somewhat wider than Davison’s intervals, as can be expected since they use less of the data. As far as we understand Stoev & Battacharya did not take truncation into account. We wonder if their methods could be modified to handle truncation.

We agree that confidence intervals are a useful way of complementing the tests performed in our paper. We also enjoyed the Stoev & Battacharya simulation-based heatmaps.

Davison next comments that his results disagree with those of Einmahl et al. (2017), who use Dutch data to conclude that there is a finite limit to the human lifespan. He writes that one possible explanation is that it is unreasonable to extrapolate from the very old persons in the Dutch data to (the even much older) supercentenarians and raises the possibility that, say, a logistic force of mortality function which first increases and then plateaus out could fit the Dutch data. Davison notes that this plateauing in fact may also show up in the Dutch data.

We completely agree with Davison’s comments: The Dutch data is dominated by ages around 100 where human mortality is clearly increasing, and to accommodate this a fitted GP distribution has to have an increasing force of mortality, or equivalently a finite endpoint. However, the (in fact quite surprising) fact shown by the IDL data, that after age 110 human mortality is at a constant plateau, is then not caught by the GP model.

Additionally, Einmahl et al. (2017) present the pooled estimates 114.1 years for the limit of life length for men and the limit 115.7 for women, based on their Dutch data: However, the IDL data and our GRG data together contain 7 men who lived longer than the limit 114.1 years, and 10 women who lived longer than the limit 115.7 years, and right now (April 19, 2018) the GRG database lists 3 women who are alive and older than 115.7 years. Jeanne Calment lived even longer than the pooled 95% upper confidence limit 120.3 for the endpoint of the lifespan for women given in Einmahl et al. (2017), and longer than the upper endpoint of more than half of the confidence intervals in this paper.

Davison also presents a quote from Einmahl et al. (2017) which argues for the use of death cohorts (rather than birth cohorts) and then in a section titled “Non-stationarity” presents an analysis which discusses the bias arising from using death cohorts. We again agree with Davison’s analysis².

5 Truncation and censoring

We appreciate the positive comments by Keiding, Davison, Stoev & Battacharya, and Zhou on our efforts to incorporate the details of the IDL sampling frame into the statistical analysis, and found

²An “extreme” example which illustrates what could happen if one uses death cohorts is as follows: Suppose that in a large country all men which are born in an even year are drafted into war and killed, and that one studies life lengths of men who died at age 110 or over in some specific even year. One conclusion would then be that male supercentenarians in this country only could live an odd number of years. This conclusion has nothing to do with the biology of aging, it is an artifact caused by the wars. Davison refers to a milder version of this example, the European heatwave of 2003 which killed many old persons.

Davison’s point process based derivation of the likelihood function for truncated data instructive.

Keiding suggests that it would be possible to use the same techniques to handle also the 2000 - mid 2003 US Data. We did not do this because the US data did not give the exact dates of deaths, only the death year, which makes it unclear how to handle the truncation. This was possible to do for the longer time period 1980 - 2000, see discussion in our paper, but seemed problematic for a 2.5 year period. As a further comment, the 2000 - mid 2003 data only include persons who were alive Jan 1, 2000, and this also had to be included in the analysis, (Rootzén and Zholud, 2016), which might make it even more fragile.

As a more general comment, taking truncation into account in the analysis often did not change estimates much. But one cannot know if this is the case or not without doing the correct analysis, which takes truncation into account. And, it did make a difference for some of the analyses.

6 Age-biased sampling and the GRG database

Zhou provides a number of examples which illustrate how conclusions may be distorted if the sample is age-biased. A general view of this is that, for all practical purposes, age bias can transform any age distribution to any other age distribution with the same, or smaller, support. The argument is as follows.

Assume that observations are i.i.d. and that the true age distribution is supported on the entire real line, and has continuous probability density function $g(x) > 0$. Further assume that the “probability” of including a life length x in the sample is $h(x)$. Then the density function of the observations in the sample is

$$e(x) = \frac{h(x)g(x)}{\int_0^\infty h(y)g(y)dy}. \quad (1)$$

Let $f(x)$ be some other probability density function on the positive real line, and assume first that there is a constant K such that $\sup\{f(x)/g(x); 0 \leq x\} \leq K$. Then, choosing $h(x) = \frac{1}{K} \frac{f(x)}{g(x)}$ in (1) gives that the density of the observations is $f(x)$. If instead $\sup\{f(x)/g(x); 0 \leq x\} = \infty$ then assume that it is possible to find an A such that $\epsilon := \int_A^\infty f(y)dy$ is arbitrarily small, and such that $\sup\{f(x)/g(x); x \leq A\} \leq K$, and set $h(x) = \frac{1}{K} \frac{f(x)}{g(x)}$ for $x \leq A$ and $h(x) = 0$ otherwise. Then equation (1) gives that the density of the observed values is

$$e(x) = \begin{cases} \frac{1}{1-\epsilon} f(x), & 0 \leq x \leq A \\ 0, & A < x, \end{cases}$$

and hence, by making ϵ small, the distribution given by $e(x)$ can be made to be arbitrarily close to the distribution given by $f(x)$. A similar argument applies if the support of $g(x)$ is a subset of the positive real line.

From a practical point of view, it is not likely that age-biased sampling would change a distribution to a substantially different one. However, age-bias could easily change an age distribution to a similar one, say, change a GP distribution to another GP distribution with a somewhat different shape parameter.

The authors of the IDL database have made a serious effort to avoid age-bias. In contrast, clicking on the link to “GRG World Supercentenarian Rankings List” on GRG (2016), and scrolling to the bottom of the page one can read “To Our Readers: Do you know of someone aged 110 or older currently living who is not on this list, but has the documents to prove it? In this case, please contact one of our two Supercentenarian Claims Investigators”. Thus GRG data are collected by investigating claims sent to the GRG group. It is inconceivable that this collection method would not lead to age-bias. Most likely it is more probable that older supercentenarians are reported. Hence the bias goes in the

opposite direction to the examples in Zhou (2018), and if anything, would change a true γ to a smaller one.

7 Ferreira & Huang analysis

The abstract of Ferreira and Huang (2018) says “We verify that negative values for the extreme value index are more likely, supporting the conclusion that models with finite endpoints seem better to fit to the human lifespan distribution”. This conclusion is based on analyses that

1. do not take truncation into account,
2. pool the IDL and GRG databases, and,
3. use yearly maxima of the data.

We think this is wrong. The reasons are as follows:

1. A main insight from the theory of statistical survival analysis is that truncation (and censoring) has to be taken into account in analysis of survival data. This in particular applies to the IDL and GRG databases, as discussed above (see also, e.g., Gampe (2010)).
2. Even disregarding the discussion above about age-bias of the GRG data, if two data sets are significantly different then it is inappropriate to merge them and consider them as i.i.d. observations of the same phenomenon. And, the IDL and GRG data sets are very significantly different. E.g., the mean of the IDL A+B data is 1.386 (0.067) and the mean of our version of the GRG data is 2.061 (0.111), and a standard t -test of equality of means gives the test quantity 5.20 and a p-value much smaller than 0.001.
3. Both the birth-year cohorts and the death-year cohorts analyzed by Ferreira & Huang contain different numbers of supercentenarians for different years. It follows that cohort maxima for different years have different distributions: the distribution of the maximum of a larger cohort is stochastically larger than the maximum of a smaller cohort. Hence analyzing cohort maxima as if they are identically distributed is wrong. This is the same mistake as made by Dong et al. (2016).

The end of Section 2 of Ferreira and Huang (2018) contains a discussion of whether truncation should be taken into account. We find this discussion confusing. It concerns the rationale for the formula

$$\frac{f(x_i)}{F(e - t_i) - F(b - t_i)}$$

on p. 724 of our paper. In this expression the numerator describes the real age distribution, which is a threshold-stable GP distribution as it should be, while the denominator comes from the sampling frame and has nothing to do with threshold stability or extreme value theory. Also, in reply to the penultimate sentence of Section 2 of Ferreira and Huang (2018): “model checking and optimizing estimation methods” is possible also for analyses which take truncation into account, see e.g. our paper and Rootzén and Zholud (2016).

8 Updated version of LATool

Section 1 of Segers (2018) mentions a private communication about how our GUI, LATool, computed the interval (b, e) which is used to correct the likelihood for truncation. However, what we told Segers

was inaccurate. For each country, our code computed b as the beginning of the year where the first death occurred, and e as the end of the year with the last death. This means that our intervals (b, e) agreed with the intervals (b, e) given in the IDL metadata, except in three cases: the b for Spain, and the e -s for Japan and the USA. We have now written an updated version of LATool, available in the supplementary material of this rejoinder, which throughout uses the intervals (b, e) given in the metadata. We have also fixed a bug in LATool and improved the estimation procedure. This has led to some changes of values in our paper. For completeness we have included updated versions of these in the supplementary material. The changes have no influence on the conclusions or discussion in our paper. However, for the new version of Fig. 5, left panel, Keiding’s comment “the observed quantiles seem to sit rather marginally among the simulations” no longer applies. We have also tried to make LATool more user-friendly, and hope it will be used for alternative analyses.

9 A misprint

Johan Segers pointed out a misprint: the paper three times says that we have excluded persons “who died in Japan after August 31, 2003” from the analysis. This should be “who died in Japan after September 30, 2004”.

10 The secret of (extremely) long life

Except for Keiding, all discussants tackle the question of whether there exists, or does not exist, a finite limit to the human lifespan, but do not write about the conclusion that there is no detectable difference between females and males, between (groups of) countries, or between time periods. The first question was also our motivation for starting this research: to find out if there is a hard biological limit to the human lifespan. However, we now think the latter conclusion is the most interesting and intriguing one. Differences would have pointed to factors which are important for long life, and which we could use to live longer – and this question interests most of us. Much of supercentenarian research is driven by this question.

A non-statistical approach to this question is taken in Jeune et al. (2010) where the authors describe the life stories of the longest-living humans. Their conclusion is “The life journeys of these very old people differed widely, and they are almost without common characteristics, aside from the fact that the overwhelming majority are women (only two are men), most smoked very little or not at all, and they had never been obese. Still, they all seem to have been powerful personalities, but decidedly not all were domineering personalities. They are living examples of the fact that it is possible to live a very long life while remaining in fairly good shape. Although these people aged slowly, all of them nonetheless became extremely frail in their final years.” This agrees completely with our result that none of the most obvious factors seem to influence the chance to live very long.

There now is quite some exciting ongoing research which tries to find genetic factors which make long life possible. And, as written by Nerman, “in the era of quick development of organ transplants, of stem cell therapies and of regenerative medicine” it seems quite possible that in the near future the human lifespan will become (much) longer. However, so far presumably these efforts have not been crowned with success – if they had we would all know about it.

So, the secret of extremely long life is still hard to find!

Acknowledgement

We thank Anthony Davison, Thomas Mikosch, and Chen Zhou for helpful comments. Research supported by the Knut and Alice Wallenberg Foundation, grant KAW 2012.0067.

Supplementary materials

Contains updated versions of [LATool](#) - the MATLAB toolbox for life length analysis, and of Figure 5 and Tables 2-5 in Rootzén and Zholud (2017). [[PDF](#)].

References

- Davison, A. C. (2018). The life of man, solitary, poore, nasty, brutish, and short. *Extremes*, 21(2):to appear.
- Dong, X., Milholland, B., and Vijg, J. (2016). Evidence for a limit to human lifespan. *Nature*, 538:257–259.
- Einmahl, J., Einmahl, J. J., and de Haan, L. (2017). Limits to human life span through extreme value theory. *Tilburg University Center for Economic Research Discussion Paper*, 2017-051.
- Ferreira, A. and Huang, F. (2018). Is human life limited or unlimited? (A discussion of the paper by Rootzén and Zholud). *Extremes*, 21(2):to appear.
- Fries, J. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of medicine*, 303(3):130–135.
- Gampe, J. (2010). Human mortality beyond age 110. In Maier, H. and et al., editors, *Supercentenarians*, chapter 13, pages 219–230. Springer-Verlag, Heidelberg.
- GRG (2016). Gereontology Research Group list of validated deceased supercentenarians. URL: www.grg.org, downloaded October 26, 2016.
- IDL (2016). International database on longevity. URL: www.supercentenarians.org.
- Jeune, B., Robine, J.-M., Young, R., Desjardins, B., Skytthe, A., and Vaupel, J. (2010). Jeanne Calment and her successors. biographical notes on the longest living humans. In Maier, H. and et al., editors, *Supercentenarians*, chapter 17, pages 285–322. Springer-Verlag, Heidelberg.
- Keiding, N. (2018). Comments to Rootzén & Zholud: Human life is unlimited – but short. *Extremes*, 21(2):to appear.
- Mayo, D. G. and Cox, D. R. (2006). Frequentist statistics as a theory of inductive inference. In *Optimality: The Second Erich L. Lehmann Symposium*. Vol. 49, Lecture Notes-Monograph Series, Institute of Mathematical Statistics.
- Nerman, O. (2018). Discussion of “Human life is unlimited – but short” by Rootzén and Zholud. *Extremes*, 21(2):to appear.
- Rootzén, H. and Zholud, D. (2016). Tail estimation for window censored processes. *Technometrics*, 258(4):95–103.
- Rootzén, H. and Zholud, D. (2017). [Human life is unlimited – but short](#). *Extremes*, 20(4):713–728.
- Segers, J. (2018). Comments on “Human life is unlimited – but short” by H. Rootzén and D. Zholud. *Extremes*, 21(2):to appear.
- Stoev, S. A. and Battacharya, S. (2018). Inference on the endpoint of human lifespan and its inherent statistical difficulty. *Extremes*, 21(2):to appear.
- Vaupel, J. W. (2010). Biodemography of human ageing. *Nature*, 464(25):536–542.
- Zhou, C. (2018). Discussion on “Human life is unlimited – but short” by Holger Rootzén and Dmitrii Zholud. *Extremes*, 21(2):to appear.