Title: Accurate age estimation in small-scale societies

Abbreviated title: A Method For accurate Age Estimation

Classification: Biological Sciences, Anthropology

Author's names: Yoan Diekmann[a], Daniel Smith[b], Pascale Gerbault[a,b], Mark Dyble[c], Abigail E. Page[b], Nikhil Chaudhary[b], Andrea Bamberg Migliano[b], Mark G. Thomas[a]

Institution: [a]Research Department of Genetics, Evolution and Environment; [b]Department of Anthropology, University College London, Gower Street, London WC1E 6BT, UK; and [c]Institute for Advanced Study in Toulouse, Allée de Brienne, Toulouse 31015, France.

Corresponding authors: Yoan Diekmann, Tel.: +44 (0)20 7679 4397, y.diekmann@ucl.ac.uk; Andrea Bamberg Migliano, Tel.: +44 (0)20 7679 7547, Fax: +44 (0)20 7679 8632, a.migliano@ucl.ac.uk; Mark G. Thomas, Tel.: +44 (0)20 7679 2286, m.thomas@ucl.ac.uk

Key words: Gibbs sampler, Bayesian age estimation, Hunter-gatherers, Fertility, Life history

ABSTRACT

Precise estimation of age is essential in evolutionary anthropology, especially to infer population age structures and understand the evolution of human life history diversity. However, in small-scale societies, such as hunter-gatherer populations, time is often not referred to in calendar years and accurate age estimation remains a challenge. We address this by proposing a Bayesian approach that accounts for age uncertainty inherent to fieldwork data. We developed a Gibbs sampling Markov chain Monte Carlo algorithm that produces posterior distributions of ages for each individual, based on a ranking order of individuals from youngest to oldest and age ranges for each individual. We first validate our method on 65 Agta foragers from the Philippines with known ages, and show that our method generates age estimates which are superior to previously published regression-based approaches. We then use data on 587 Agta collected during recent fieldwork to demonstrate how multiple partial age ranks coming for multiple camps of hunter-gatherers can be integrated. Lastly, we exemplify how the distributions generated by our method can be used to estimate important demographic parameters in small-scale societies, here age-specific fertility patterns. Our flexible Bayesian approach will be especially useful to improve cross-cultural life history datasets for small-scale societies for which reliable age records are difficult to acquire.

SIGNIFICANCE STATEMENT

Understanding demographic and evolutionary processes shaping human life history diversity depends on precise age estimations. This is a challenge in small-scale societies, and especially those who do not follow a calendar year. Our method opens new possibilities in demographic and life history studies allowing for cross-sectional data to be incorporated in cross-cultural comparisons and for a better understanding of the adaptive importance of human life history variation.

\body

INTRODUCTION

Accurate estimation of the age of individuals is essential in evolutionary anthropology. Major questions in the field require an accurate inference of the timing of life history events, such as age at menarche, age at first reproduction, age at cessation of reproduction, inter-birth intervals and death. Age is also essential when assessing infant growth, developmental trajectories and estimating age structure properties of a population (e.g. the potential for population growth or decline, recovering signatures of epidemics and assessing vulnerability to ecological perturbations). Humans have important derived life history features, such as shorter inter-birth intervals, longer lifespan, extended post-reproductive longevity and childhood dependence (1). These life history traits vary across species in the slow-fast continuum (2), and they likely vary within humans in response to differences in ecology such as differential mortality rates (3) and energetics (4). However, due to unreliable age estimates, very few studies have highlighted variability in life history traits in traditional societies (3, 5-6). The challenge of estimating ages is particularly problematic for populations where individuals do not relate their age to calendar years, as is the case among many hunter-gatherer and other small-scale societies (7-8). Although longitudinal studies are an ideal approach to address questions about variation in life history traits in small-scale populations, these are rare (although see 7, 9). There is consequently a need for methods to estimate ages based on cross-sectional data from these populations.

A few approaches have been proposed to estimate ages in small-scale societies (reviewed in 7). The simplest one is visual inspection and approximate clustering into age cohorts (e.g., infant, child, teen, adult, old age). A clear disadvantage of this method is its lack of precision, as establishing life history strategies requires a refined age structure. Furthermore, differences in physical appearance trajectories in forager populations in comparison to western counterparts are likely to cause misattribution of ages. For instance, forager infants are often small and underdeveloped, appearing younger than their western peers, while older individuals may appear older when compared to western individuals of the same age. An alternative class of approaches are indirect demographic models developed in the field of human demography (21), which are characterized by model parameters that are estimated based on actual population data. For example, Howell

(8) applied a ‘steady-state model’ approach to the Dobe !Kung foragers. This method assumes a stable population structure, ascertains a relative age list of all individuals and estimates the death and fertility rates of the population. This approach permits an approximation of the population age structure by mapping these rates onto different life-tables (in which, for example, 80% live to age 1, 75% live to age 2, etc.) and selecting the life-table with the best correspondence. Given that these life tables are created from very different populations, caveats of this approach include the difficulty to find matching life tables, particularly for growing or declining populations for which these rates are unknown (7). Crucially, stable population models fix the proportion of individuals that live up to a certain age, which may obscure differences in life history adaptations and demography.

To overcome these problems, Hill and Hurtado (7, but also see 10 for the Hadza) designed an alternative method to estimate the ages of Ache hunter-gatherers that did not assume a stable population. It is based on a relative age list including all individuals, with absolute ages for a subset of individuals. The relative age list was constructed by first dividing the population into age cohorts containing individuals of approximately the same age. Each individual ranked all others within their cohort, as well as those in the cohorts above and below them (i.e. either older or younger than themselves). These relative lists were combined into cohorts and then a master population list by selecting the rankings with the smaller number of contradictions. The absolute ages of some of the individuals were obtained from birth certificates, estimated from known events, or by an “age-difference chain” (individuals were questioned about their age at the time a younger individual was born by picking an individual of a known age and matching their age at the time of birth of the younger individual, 7). Given these absolute ages and the relative age list, a fifth-order polynomial curve was fitted with relative age rank as independent and age as dependent variables. Finally, the ages of the remaining individuals were estimated as the value of the polynomial curve at the corresponding rank.

Despite improving upon previous methods, this approach still presents several drawbacks. First, the choice of fifth-order polynomial is arbitrary. Previous authors (10) have, for example, used third-order polynomials. Some ages may be fitted poorly by a polynomial, while overfitting may also be an issue, especially for data sets with few known ages. In addition, the uncertainty associated with any age

estimate is not taken into account. This is particularly problematic in age-difference chains as the error is cumulative, leading to high uncertainty, especially for older individuals. For example, for the Ache with known ages Hill and Hurtado (7) have shown that the error in age estimates using this age-difference chain is approximately +0.5 years (SD=1.2) for each 12.5 year interval. Although relatively small, with the oldest individuals potentially overestimated in age by an average of ~2 years, this method does not control for the concomitant increase in error with age. Based on standard deviations this would be between +8.6 and -3.4 years for the oldest individuals based on their predicted age from the regression model. This has particular relevance to the estimation of age at some important life history events in later life, such as age at last reproduction and menopause.

Here, we present a new Bayesian method for age estimation improving upon previous approaches. Bayesian approaches have previously been designed and successfully applied in, for example, Paleodemography (19-20) and Radiocarbon dating in Archaeology (22), however, they are not readily applied to data typically collected in anthropological fieldwork on small-scale societies. Our method requires two inputs. First, a single ranking or multiple partial rankings of individuals by age obtained from interviewing members of the population. And second, an arbitrary *a priori* age distribution per individual chosen by the researchers familiar with the population, that can be either based on accurate measures or 'eyeballing'. These two pieces of information are combined using a statistical inference technique called Gibbs sampling, generating a posterior age distribution for each individual. This posterior distributions represent all that can be known about the ages of that individual given the age ranks and prior age distributions. We show that our method generates more accurate age estimates than regression-based approaches on 65 individuals from a hunter-gatherer society with known ages. As further empirical validation, and to show the flexibility of our method for actual fieldwork, we present a case study on Agta foragers from Palanan, the Philippines. Finally, we analyse age-specific fertility patterns in the Agta, fully integrating the uncertainties in the estimated ages of mother and offspring. This demonstrates how the posterior distributions produced by our method can be reliably used for estimating important demographic parameters in small-scale societies for which precise dates of birth do not exist. Our method opens new possibilities in demographic and life history

studies, allowing for cross-sectional data to be incorporated in cross-cultural comparisons.

RESULTS

Validation and benchmarking: Bayesian out-performs regression-based approaches to age estimation

First, we assess how well our Bayesian approach estimates ages compared to regression methods. We apply five-fold cross validation (CV), i.e. we randomly partition 65 Agta with known ages (obtained from reference 13) into five groups of 13 individuals, consider the ages of the individuals in each group in turn as known, and estimate the ages of the remaining individuals. For each of the five partitions, this procedure yields 52 estimates that are then compared with the true known ages. See Materials and Methods for details. The results are summarised in Figure 1 and SI Appendix, Sup. Table S1. The distribution of differences between known age and mean age estimated by our method across all five CV partitions shows that the median error of the differences per individual is about 0.29 years (i.e. four months), and the mean 0.91 years (i.e. 11 months). Estimation accuracy becomes worse for older individuals, whose ages are inherently more difficult to estimate due to wider prior age brackets and larger age differences between the individuals (see SI Appendix, Sup. Table S1). Interestingly, similar results are achieved even when no age is considered known and ages are estimated based on rank and age brackets alone. The near-equivalence of the Bayesian method with and without known ages is also supported by statistical comparison of the two distributions of error: a Bayesian *t*-test finds no evidence for different means, while a non-parametric two-sided Kolmogorov-Smirnov (KS) test reports no significant differences between the two distributions (Bayes factor (BF) = 0.23, $p = 0.61$, see Figure 1).

In comparison to the Bayesian approach, polynomial regression has a higher median error of the differences per individual of around 1.16 years (14 months) and a high mean of 2.66 years (32 months). The latter is the result of multiple outliers in the error distribution caused by high estimation errors for very young or very old individuals, especially when the closest individual with known age is far from these individuals. For example, the first individual with known age in partition three (see

Figure 2) has rank 12. Counter-intuitively, regressing on both known ages and midpoints of the age brackets does not improve the estimation (see Material and Methods for the rationale behind including the midpoints). The mean error for polynomial regression fitted with midpoints of the age brackets is 52 months, and comparing it to the distribution without midpoints via a Bayesian *t*-test and KS test yield very strong evidence for greater error in the model using midpoints (BF > $4x10^{20}$, $p < 1x10^{-10}$, see Figure 1). We also tested a third approach based on local regression (LOESS, 17), which drops the requirement for the data to fit a fifth-order polynomial and allows for more flexible curves. LOESS shows intermediate performance with a median error of 0.64 years (seven months). See SI Appendix, Sup. Table S2 for *p*-values and Bayes factors of all pairwise comparisons of error distributions, including LOESS.

We tested the influence of the number of known ages, employing two to 13-fold cross validation. The performance of the Bayesian approach is not significantly influenced by the number of known ages. This is not the case for the polynomial regression, for which large differences are observed, especially when fewer ages are known, mostly reducing the accuracy (see SI Appendix for details). Furthermore, we asked how robust the approaches are to errors in known ages and ranking order. SI Appendix, Sup. Figure S4 shows that our Bayesian method is not influenced by slight errors in known ages, whereas polynomial regression and to a lesser extent LOESS follow a trend towards worse performance. Errors in ranking order cannot be tested independently of errors in known ages for regression approaches. We therefore only assessed our Bayesian approach, and find a clear impact of errors in ranking order on the estimation accuracy as shown in SI Appendix, Sup. Figure S5. Yet, even with 40% errors in the ranking order the estimation accuracy is comparable to that of polynomial regression when supplying the correct order. Lastly, we explored how well the resulting posterior distributions quantify the estimation uncertainty. To be useful as quantification, a 95% credible interval for example should contain the true age in 95% of the individuals whose age is being estimated, while a 50% credible interval only in half of the individuals. We tested this with highest posterior densities (HPD), and confirmed that HPDs closely mirror estimation uncertainty (see SI Appendix, Sup. Figure S2).

In summary, we observe that our Bayesian approach outperforms both LOESS and polynomial regression. It achieves this accuracy nearly independently of the availability of known ages and correctly quantifies estimation uncertainty. Lastly, it is robust to errors in known ages and to some extent in rank order.

Palanan Agta Case study: A flexible method for fieldwork data

After testing the data in a longitudinal dataset with known ages, we applied our aging methodology to an anthropological cross-sectional case study on Agta foragers from the Philippines, for whom most ages are unknown. In particular, we highlight two key aspects of our approach: first, the flexibility of our method in dealing with fieldwork data by allowing for multiple partial ranks in age estimation; and second, exemplifying how the uncertainties in age estimates can be integrated into subsequent analyses, such as estimating age-specific fertility patterns, which requires the estimation of both mothers and child ages, potentially increasing estimation errors.

A key difficulty with small-scale societies – including the Agta – is that individuals living in geographically distant camps rarely know each other well enough to accurately rank each other's ages. As a result, this loose pattern of familiarity among individuals precludes the assembly of a single age rank. Rather, multiple partial ranks are generated, in our case 266 partial ranks, that include different – yet overlapping – subsets of individuals, but never the entire population. One of the great flexibilities of our Bayesian approach is that this situation can intuitively be accommodated. We present our approach to multiple partial ranks informally here, and give more details in the SI Appendix. In the first step, consistent partial ranks are merged. For example, $(\mathrm{A}, \mathrm{B}, \mathrm{C})$ and $(\mathrm{B}, \mathrm{C}, \mathrm{D})$ are consistent and can be merged to yield $(\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D})$. In contrast, $(\mathrm{A}, \mathrm{B}, \mathrm{C})$ and $(\mathrm{B}, \mathrm{A}, \mathrm{D})$ are not consistent and therefore kept separate. Longer ranking orders that result from merging tend to impose stronger constraints on the prior age distributions, especially for individuals otherwise at either end of the partial rank, which results in narrower posterior distributions and consequently more accurate age estimates. Together with the priors on the individuals' ages, all partial ranks resulting from this merging step are then used as input for separate runs of the Gibbs sampler, where a run produces distributions of

ages for each individual contained in the partial rank. At this stage one has multiple results from independent applications of our Bayesian approach to different partial ranks and the same age priors. The last step is to merge all distributions that belong to the same individual, generating a final age distribution per individual. To this end, the different distributions are combined to form a weighted mixture density, which can be thought of as simply adding up the various distributions and rescaling them to integrate into one. This way, ranking orders that have been frequently reported by multiple individuals are naturally weighted more than those reported once or only a few times. The upper two panels in Figure 3 demonstrates this procedure for two Agta, where combining the age distributions for a mother and her child yields a distribution for the age of the mother at the time of parturition.

Besides its flexibility to deal with multiple partial ranks, a distinctive feature of the Bayesian approach presented here is that it produces full posterior age distributions that quantify uncertainty rather than mere point estimates. Figures 3 and 4 illustrate how the full information in the posterior can be integrated into subsequent analyses, here age-specific fertility. Computing the age at parturition is trivial when the age of both mother and child are known exactly: simply calculate the difference. However, if the age of the mother, the child, or both are uncertain and therefore described by a distribution, the solution becomes less obvious. This is precisely the case here, as our age estimation procedure results in distributions that capture the uncertainty in the age estimate. In Figure 3, we use convolution to derive the distribution of age at parturition for a mother (see Materials and Methods for the definition of convolution), which explicitly considers the uncertainty about maternal and child ages. This analysis was performed on all mother and child pairs, forming the mixture of the resulting distributions (think of as ‘averaged’, i.e. stacked and normalised) in order to obtain the overall distribution of the age at parturition in the Agta population. Figure 4 depicts this posterior distribution of age at parturition separately for cases where both the mother’s and the child’s ages are known exactly from birth certificates (histogram) and for all other cases (density curve). While we do not necessarily expect the distributions to be the same as fewer precise ages are available for older individuals (see SI Appendix, Sup. Table S3), we nonetheless fail to reject the null hypothesis that both are sampled from the same distribution (KS-

test $p > 0.10$). We interpreted this as an internal check validating our approach and results.

DISCUSSION

This study introduced a Bayesian approach to estimate ages in a fully probabilistic framework. Its strengths are high accuracy and great flexibility. Initial age ranges or prior distributions can be chosen from a wide spectrum of distributions to reflect the level of confidence in the *a priori* age estimate for each individual: from point masses when date of birth is known, to wide uniform distributions when ages are vaguely estimated and lie in a poorly informed range. The second type of input data that is required is a ranking of individuals by age. However, our approach can also work on multiple partial ranks, a common difficulty when aging small-scale societies. Figure 5B exemplifies how these two data types are integrated to produce posterior distributions that fully capture and quantify the uncertainty in the resulting age estimates.

By comparing our method to regression-based approaches, we have demonstrated that the Bayesian approach outperforms all regression methods considered, and furthermore correctly quantifies estimation uncertainty. Notably, this is true even when no known dates of birth are provided; a situation where regression based approaches cannot be applied. Hence, our approach can also work when absolute ages for all or most individuals are not available. However, we caution that the number of individuals and the density with which they cover the range of ages is critical for the accuracy of estimated ages. No accurate estimate is possible with only a few individuals of very different ages. Nonetheless, the necessary data can be obtained in short field trips, which should make age estimates for various small-scale societies readily available, facilitating future studies on the evolution of human adaptive variation.

The Agta case study demonstrated that our method performs well in typical fieldwork conditions and challenges. The large geographical area of Agta camps made it impossible to compile a single complete age rank, and therefore, we extended our basic Bayesian framework to deal with partial ranks (see Figure 3). This demonstrates that specific social organisations with particular traits can be

integrated by relatively simple extensions of our approach, making our method widely applicable in diverse fieldwork conditions.

Finally, we analysed the results we generated for the Agta to illustrate how the posterior age distributions produced by our method can be used in subsequent analysis. Age-specific fertility patterns are a fundamental aspect of population structure and are necessary to understand demographic and model population processes (11). Figures 3 and 4 show how the uncertainties in the posterior age estimates can be propagated through the different steps of the analysis and integrated into the final result. In contrast, approaches based on summary statistics (e.g. mean and median, that by definition do not capture the full information contained in the data), or binning point estimates into arbitrary age classes, may distort and inflate confidence in final results, and do not allow comparisons at the individual level.

The example above illustrates the importance for future work to derive statistical methods that use the posterior age distributions directly and therefore the full information content of the data. In these cases, the potential of our probabilistic approach can be fully reached, although we show in Figure 1 that point estimates (mean age of the posterior distribution) generated by our method already improve accuracy. Even though no generic solutions exist for analyses involving ages, standard approaches such as resampling from the posterior distribution can be implemented on top of the output produced by our method. It should be noted that, as with all MCMC-based Bayesian approaches, the MCMC chain, once mixed, is a sample from the posterior, making such approaches easy to implement.

In summary, our Bayesian approach has the potential to increase the utility of cross-cultural life history datasets for hunter-gatherers and small-scale societies living in various environments, and enable robust and powerful statistical comparisons between human population groups to shed light on the adaptive processes shaping variability in human life history.

MATERIALS AND METHODS

Bayesian estimation of ages

In contrast to previous approaches, we address age estimation in a fully probabilistic framework. For a set of individuals, two types of input data are required: (i) a ranking or ordering of all individuals by age of the type *A* is younger than *B* is younger than *C* etc.; and (ii) an *a priori* age distribution per individual. For example, in the simplest case the *a priori* distributions may be uniform, i.e. given by hard bounds on the plausible age of the individual of the type not younger than l and not older than u with all ages in between equally probable. We also refer to the interval $[l, u]$ as age bracket. We require rank order and age brackets to be compatible, that is a combination of ages must exist that has non-zero prior probability and satisfies the ranking order. Note that we relax the requirement of a single ranking including all individuals in the main text to allow for multiple partial rankings. Ranking and prior age distributions are processed to generate a probability distribution of age per individual. If an individual is not included in any ranking order, the *a priori* age distribution and age bracket is all that can be known about the individual's age.

In the following, we describe how these age distributions are generated by Gibbs sampling, while the mathematical definitions can be found in the SI Appendix. The heart of the procedure is iterative sampling of random numbers, which are constrained in a way to gradually approach the desired age distributions. Convergence to the correct distribution is certain and can be mathematically proven. As an example, panel A of Figure 5 illustrates the initialization and 2 sampling steps for five hypothetical individuals. Say the ranking of the individuals is reflected by their label, i.e. 1 is younger than 2 is younger than 3 etc., and their ages have been bounded *a priori* as shown by the age brackets. As a starting point for the sampling, we initialize the age of each individual to be the smallest possible value that satisfies both the constraints imposed by the ranking and the age brackets. In our example, that is achieved by choosing the left bound of the age bracket for individuals 1, 2 and 3, however, individuals 4 and 5 must be older than individual 3 and therefore appear in immediate succession after individual 3. Note that this is only one of many possible starting configurations, but as long as the ordering and age range constraints are satisfied the actual starting point is irrelevant and all yield equal results. After setting the initial values, each individual is considered in turn from the youngest to the oldest and assigned a new age by random sampling. The essential requirement for Gibbs sampling to work is that the ranking constraints and age

brackets are not violated. This means that an appropriate range to sample a new age from has to be chosen at each step, in panel A of Figure 5 for example marked by grey shading, which can be derived as follows. The youngest possible age is the higher value out of the preceding individual's sampled age and the lower bound of the current individual's age bracket. The oldest possible age is the lowest value out of the following: the upper bounds of the current individual, the upper bound of all succeeding individuals, and the next individual's age sampled in the previous iteration. If sampling is repeated often enough, this procedure results in individual age distributions that combine both the information contained in the age brackets and the age ranking. For the individuals introduced in Panel A of Figure 5 and uniform prior distributions, the effect is shown in Panel B. Intuitively, one can think of the age ranking information as "distorting" the prior distributions. Note that the approach accommodates arbitrarily small age brackets, in the extreme even containing only a single value. Hence, if the age of certain individuals is known with certainty, this information is fully used without any change to the sampling scheme described above.

The results represent all that is known about the age of the individuals, and are a combination of all the information already contained in the input; no information has been discarded or added based on additional assumptions. This also implies that if the age brackets or the ranking contain errors, so will the output of our method. However, as we show in the Results section, we are able to extend our method to work with multiple partial ranks, which allows us to avoid making choices and potentially introduce ranking errors in cases where rank order is unclear. The fundamental advantage of our method is that its output is a distribution. This allows subsequent analyses to incorporate the full uncertainty associated with point estimates (e.g. by confidence intervals around the mean age), or in the best case to directly use the full age distribution of an individual (see for example age at parturition estimation below) and therefore the entirety of the available information.

Validation and benchmarking

We validate our approach on 65 Agta hunter-gatherers from Casiguran (the Philippines), whose exact dates of birth are known (13), and can be directly

compared to the estimates generated by our Bayesian approach. Ideally, we would have validated our method on multiple samples from populations with different age structures. However, we are not aware of any other public dataset providing both pictures and exact ages that we require to run our method. As with any validation, we therefore caution that our performance results do not necessarily generalize beyond the dataset we used. Yet, as we do not make any assumption about the population, including a specific age structure, we are confident that the performance results we present extrapolate well.

As input data, we derived a relative ranking from the known dates of birth, and three of the authors (DS, AEP & MD) assigned upper and lower age bounds to these individuals based solely on visual inspection of the accompanying pictures (done prior to knowing the actual dates of birth). As photographs were taken in different years (between 1972 and 2010), all ages and age estimates were adjusted to the present day (2015), hence the youngest age is 15 and the oldest is 93. In order to make the results comparable, we summarized each posterior distribution by its mean, which can then be easily compared to the known age of the individual by calculating the difference between the two.

Besides validating our results against the known true ages, we also compare the quality of our inference against two alternative methods: the regression approach, fitting a fifth-order polynomial (7), and a non-parametric alternative based on local regression with LOESS (14).

We implement a five-fold cross validation strategy: We randomly split the data into five groups of 13 individuals and consider each group in turn. For each group, we estimate the regression equation and use it to deduce the ages of the remaining individuals. Within the Bayesian framework, known ages are taken into account by choosing discrete probability masses as priors for the age of an individual rather than uniform densities over an age interval. Figure 2 sums up our setup: the random partitioning of the individuals in five groups (top row), the known ages and the lower and upper limits (i.e. age brackets) derived from the individuals' pictures, and the regression curves. The lower and upper age limits vary between individuals, with older individuals tending to have wider ranges as their age is generally associated with more uncertainty. Note that the regression approaches do not accommodate information on the age ranges provided by the age brackets, whereas our Bayesian

approach does. We therefore also test a fifth-order polynomial regression fitted not only on the known ages of 13 individuals for a given cross validation partition but on the middle values of the age brackets for all other individuals as well. As far as the differences between the method presented here and regression allow, this ensures a fair comparison as both approaches are provided with equivalent input. Lastly, in order to test how our method would work in a situation where exact ages are impossible to obtain, we also apply our approach entirely without known ages, i.e. solely relying on the information from the age brackets and the ranking of individuals.

Case study: Palanan Agta

We apply our age estimation method to data we collected on the Palanan Agta, a hunter-gatherer population from north-east Luzon, north of the Casiguran Agta, in order to demonstrate the application and flexibility of our method. We give a detailed description of the collection procedure we devised for the two types of data required as input—the ranking orders, and the age brackets for all individuals—in the Sup. Materials and Methods section of the SI Appendix. Ethical approval for this project was granted by the University College London Ethics Committee (UCL Ethics code 3086/003) and carried out with permission from local government and tribal leaders in Palanan. Informed consent was obtained from all participants, and parents signed the informed consents for their children (after group and individual consultation and explanation of the research objectives in the Agta language).

Estimated age at parturition based on age distributions of mother and child

Let the age of mother and child be modelled by random variables M and C, respectively. Analogous to the case where ages are known exactly, the age at parturition—say P—is then described by the difference between the two random variables, $P = M - C$. As M and C are both defined by distributions, so is P, and the full probabilistic description of the age at parturition we seek is given by the probability density function (pdf) of P, say $f_P(x)$. It can be derived from the pdfs of M and C by a mathematical operation called ‘convolution’: let $f_M(x)$ and $f_C(x)$ be the pdfs of M and C, respectively, then

$$f_P(x) = \int_{-\infty}^{\infty} f_M(\tau)\, f_C(\tau - x)\, d\tau.$$

Convolution can therefore be thought of as an operation transforming two distributions into one, as illustrated in Figure 3.

Implementation and statistical analyses

The Gibbs sampler has been implemented in Python 2.7 (15) and can be downloaded from our website at http://www.ucl.ac.uk/mace-lab/resources/software. See SI Appendix for detailed information including burn-in, thinning and various diagnostic statistics.

All analyses and plotting were implemented in the statistical analysis programming language R version 3.1.3 (16). Regression analyses were performed using the functions 'lm' (17, ch. 4) and 'loess' (17, ch. 8), Kolmogorov-Smirnov (KS) statistical tests with 'ks.test' and convolution with the function 'convolve' all from the R library 'stats'. Bayesian *t*-tests were computed by the function 'ttestBF' (23) from the 'BayesFactor' library. The KS-test in Figure 4 is performed by rejecting the null hypothesis at level α if the KS-statistic $D_{n,n'}$ is greater than the critical value approximated by $c(\alpha)\left(\frac{n+n'}{nn'}\right)^{1/2}$, with $c(0.1) = 1.22$ (see Tables 54-55 in 18) and n and n' being the sample sizes, here 23 exact ages at parturition (summarized in the histogram) versus a distribution derived from 324 mother/child pairs.

ACKNOWLEDGEMENTS

We thank the four referees and the editor for the many valuable comments that greatly improved the manuscript.

M.G.T. and Y.D. are supported by a Wellcome Trust Senior Research Fellowship (Grant 100719/Z/12/Z, "Human adaptation to changing diet and infectious disease loads, from the origins of agriculture to the present" awarded to M.G.T.). D.S., P.G., N.C., M.D., A.E.P., A.B.M., and M.G.T. are supported by Leverhulme Programme grant RP2011-R-045 to A.B.M. and M.G.T. M.D. is also supported by ANR Labex IAST.

A.B.M. and M.G.T. conceived the project. A.B.M, A.E.P., N.C., D.S., & M.D. developed the data collection method, D.S., M.D., A.E.P. collected the data. Y.D., P.G., and M.G.T. initiated the analysis approach. Y.D. implemented and performed the analyses. Y.D., D.S., P.G., N.C., A.B.M., and M.G.T. wrote the manuscript with the help from all authors. All authors contributed substantially to revisions and gave final approval for publication.

REFERENCES

1. Kaplan H, Hill J, Lancaster J, Hurtado A M, Hill K I M, Lancaster J, Hurtado A M (2000) A theory of human life history evolution: diet, intelligence, and longevity. *Evolutionary Anthropology 9*:156–185.
2. Charnov E (1993) *Life history invariants: some explorations of symmetry in evolutionary ecology*. New York City: Oxford University Press.
3. Migliano A B, Vinicius L, Lahr M M (2007) Life history trade-offs explain the evolution of human pygmies. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(51):20216–20219.
4. Stearns S (1992) *The evolution of life histories* (Oxford University Press, Oxford).
5. Rozzi F V R, Koudou Y, Froment A, Le Bouc Y, Botton J (2015) Growth pattern from birth to adulthood in African pygmies of known age. *Nature Communications 6*:7672.
6. Walker R, Gurven M, Hill K, Migliano A, Chagnon N, De Souza R, Djurovic G, Hames R, Hurtado A M, Kaplan H, Kramer K, Oliver W J, Valeggia C, Yamauchi, T (2006) Growth rates and life histories in twenty-two small-scale societies. *American Journal of Human Biology 18*:295–311.
7. Hill K, Hurtado A M (1996) *Aché Life History: The Ecology and Demography of a Foraging People* (Aldine de Gruyter, New Haven).
8. Howell N (1979) *Demography of the Dobe !Kung* (Aldine, London).
9. Early J D, Headland T N (1998) *Population Dynamics of a Philippine Rain Forest People: The San Ildefonso Agta* (University Press of Florida, Gainesville).
10. Blurton Jones N G, Smith L C, O'Connell J F, Hawkes K, Kamuzora C L (1992) Demography of the Hadza, an increasing and high density population of Savanna foragers. *American Journal of Physical Anthropology 89*(2):159–181.
11. Weiss K M (1973) A Method for Approximating Age-specific Fertility in the Construction of Life Tables for Anthropological Populations. *Human Biology 45*(2):195–210.
12. Walsh B (2004) Markov Chain Monte Carlo and Gibbs Sampling. *Lecture Notes*

for EEB:1–24.

13. Headland T N, Headland J D, Uehara R T (2011) *Agta Demographic Database: Chronicle of a hunter-gatherer community in transition* (SIL Language and Culture Documentation and Description, 2).
14. Cleveland W S (1979) Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association, 74*(368):829–836.
15. Python Software Foundation. (2016) Python Language Reference.
16. R Core Team (2012) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/
17. Cleveland W S, Grosse E, Shyu W M (1992) Local Regression Models. *Statistical Models in S,* eds Chambers J M, Hastie J J (Wadsworth & Brooks, Pacific Grove, CA), pp 309–376.
18. Pearson E S, Hartley H O, eds. (1972) Biometrika Tables for Statisticians, Vol. 2 (Cambridge University Press), pp 117–123.
19. Caussinus H and Courgeau D (2010) Estimating age without measuring it: A new method in paleodemography. Population-E, 65(1):117-145.
20. Séguy I, Caussinus H, Courgeau D and Buchet L (2013) Estimating the age structure of a buried adult population: a new statistical approach applied to archaeological digs in France. American Journal of Physical Anthropology, 150(2): 170-183.
21. United Nations (1983) Manual X: Indirect Techniques for Demographic Estimation. Population Studies series, vol. 81. New York: UN Population Division.
22. Ramsey CB (2009) Bayesian Analysis of Radiocarbon Dates. *Radiocarbon*, *51*(1):337–360.
23. Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G (2009) Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review 16*(2):225–237.

FIGURE LEGENDS

Figure 1. *Validation and benchmarking of the Bayesian approach*. Boxplots of absolute differences between estimated and known ages for all 65 individuals in all five partitions from the Headland database of Agta (13) (see also Figure 2), for five different methods of estimation, i.e. fifth-order polynomial, fifth-order polynomial with mid-point age estimates, LOESS, Bayesian posteriors approximated by the mean, Bayesian posteriors computed without taking into account known ages approximated by the mean. Note that the latter distribution without known ages only comprises 65 differences as no multiple partitions exist. Statistical comparisons are performed with Bayesian *t*-tests quantifying the strength of evidence for different means of the logged distributions via Bayes Factors (BF; BF greater than three are considered positive evidence, above 150 as strong evidence) and two-sided non-parametrical Kolmogorov-Smirnov tests assessing difference between distributions (see SI Appendix, Table S2 for all pairwise comparisons). The y-axis is in log-scale to highlight the majority of differences that are below 10; see SI Appendix, Figure S6 for the raw values.

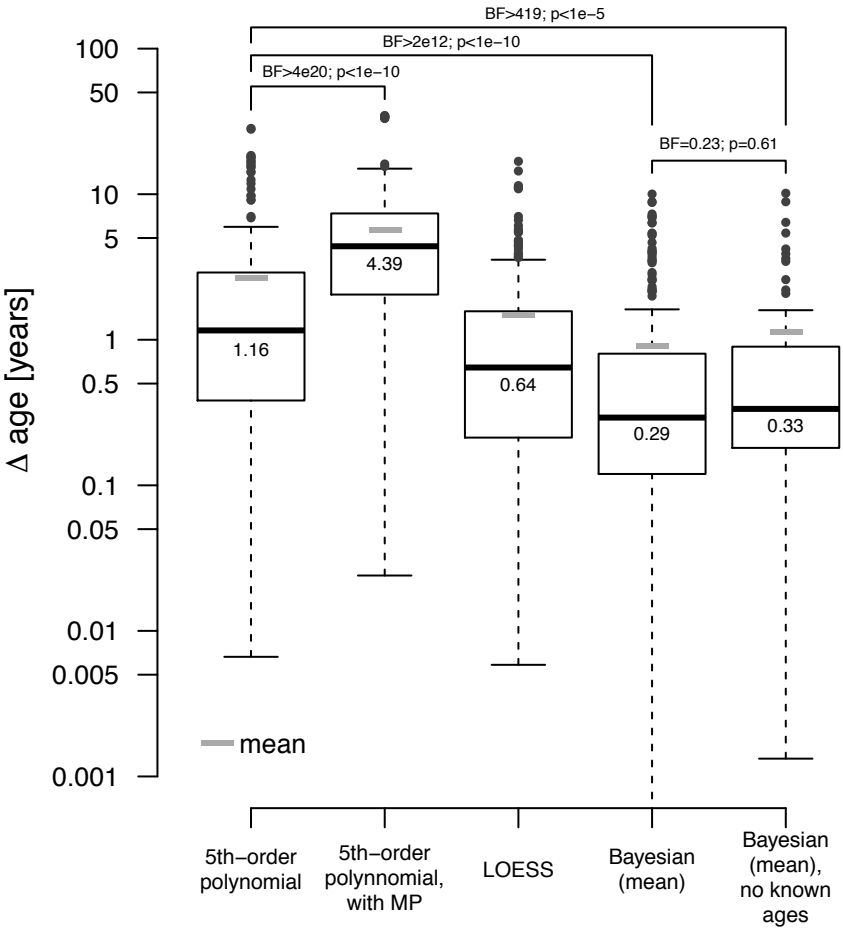
Figure 2. *Experimental Setup and results of validation and benchmark of the Bayesian approach*. We show the results of four different ways to estimate ages, including the Bayesian approach presented here. We performed five-fold cross validation; that is randomly partitioned the 65 Agta with known ages in the Headland database (13) into five groups of 13 individuals each (groups given at the top of the first panel), and used each group as the basis to estimate the age of the remaining individuals. Each panel shows the results for the five partitions, from top to bottom: fifth-order polynomials, fifth-order polynomials fitted on 13 known ages and midpoints of the age brackets for the remaining individuals (age brackets are the lower and upper age limits, inferred by the authors' from photographs of the individuals), LOESS (17), and finally the Bayesian method, including the results of a sixth run where no ages are considered known.

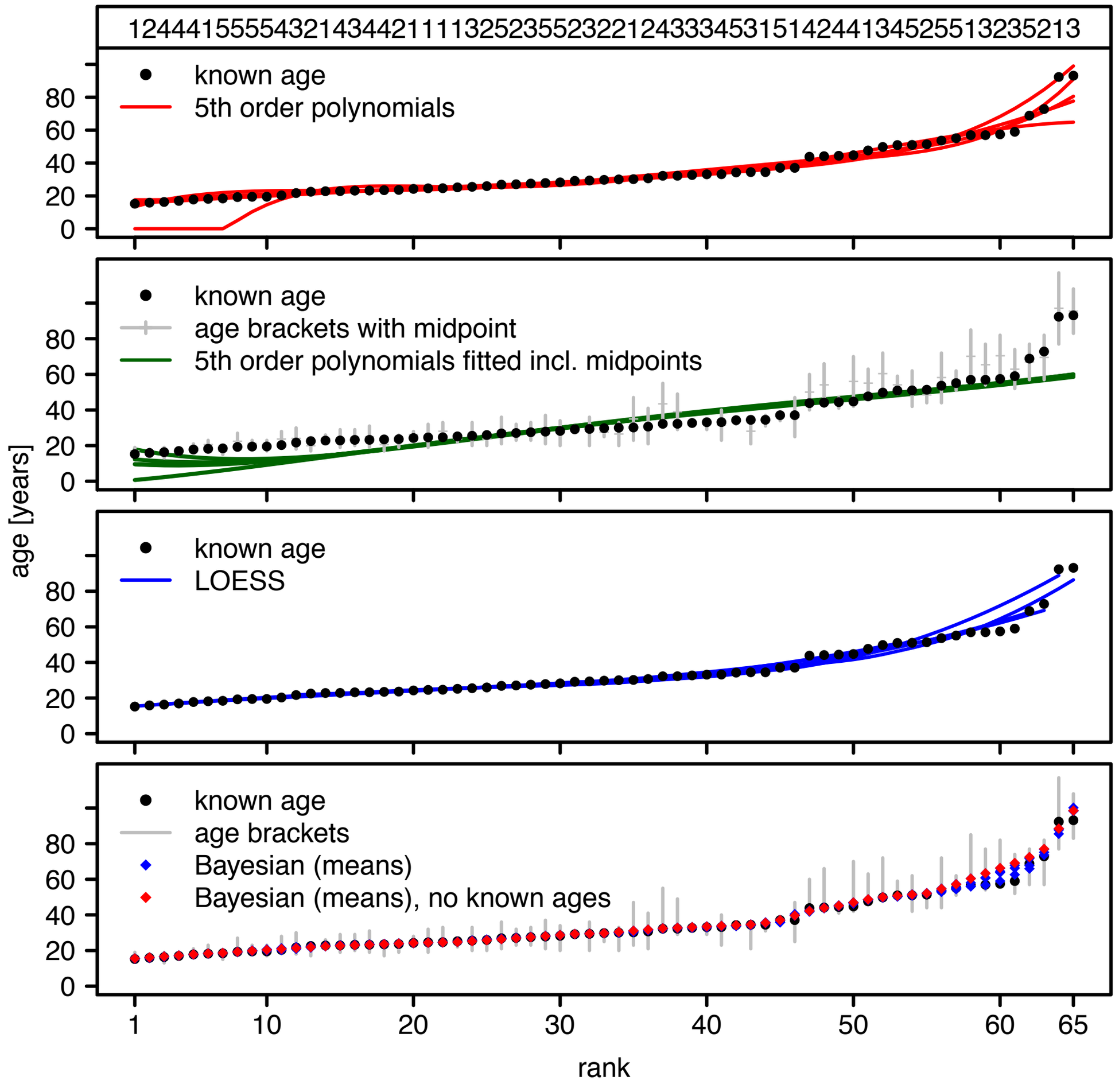
Figure 3. *Integrating uncertainties to estimate the mother's age at parturition*. The upper two panels illustrate how distinct partial rankings of individuals are

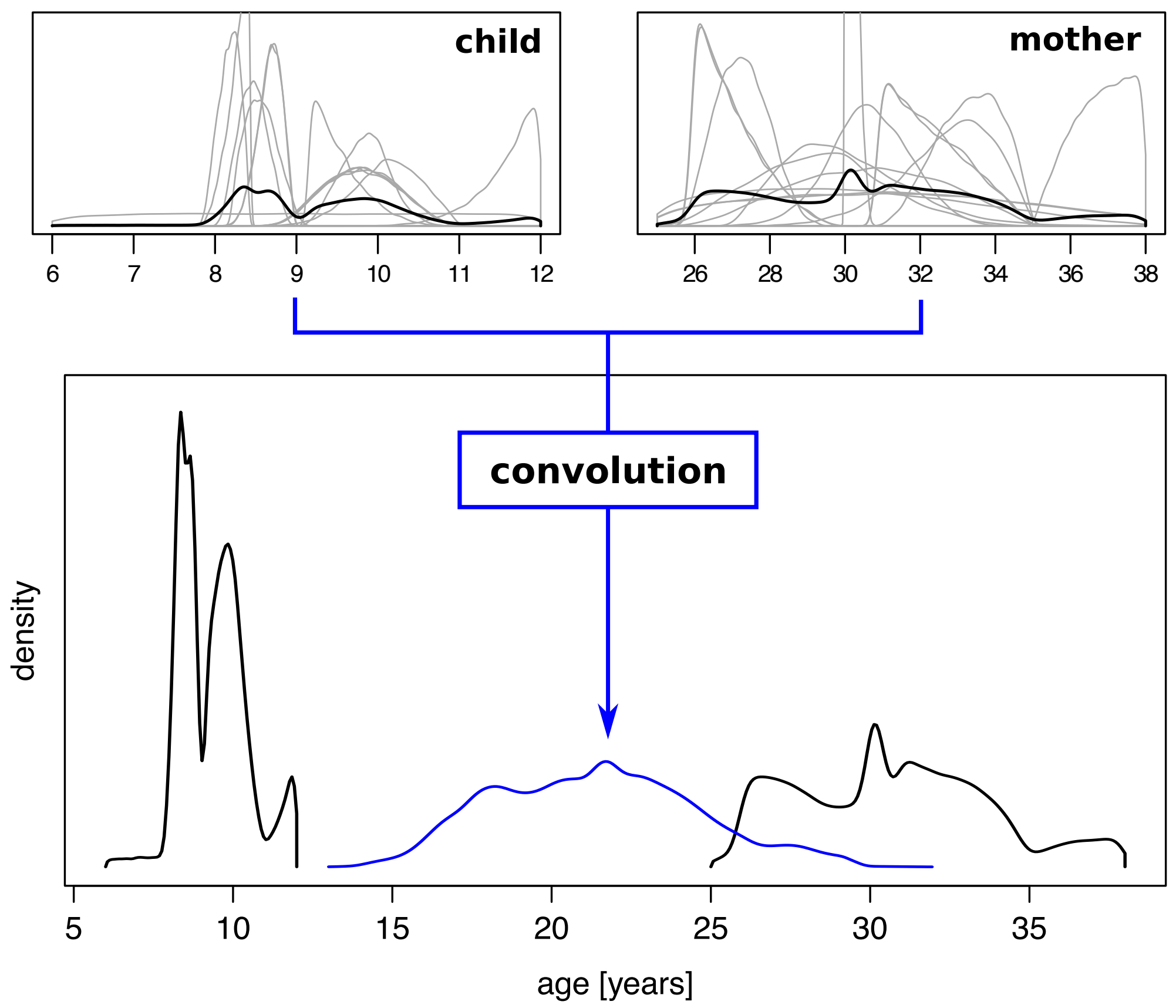
combined by averaging the resulting age distributions (grey density curves) to give an overall age distribution (black density curves) per individual. The pair of individuals was chosen to be mother (right upper panel and right distribution in lower panel) and child (left upper panel and left distribution in lower panel), allowing us to "convolve" (see Materials and Methods) the age distributions and obtain the posterior distribution of the mother's age (lower panel, blue density curve) at parturition.

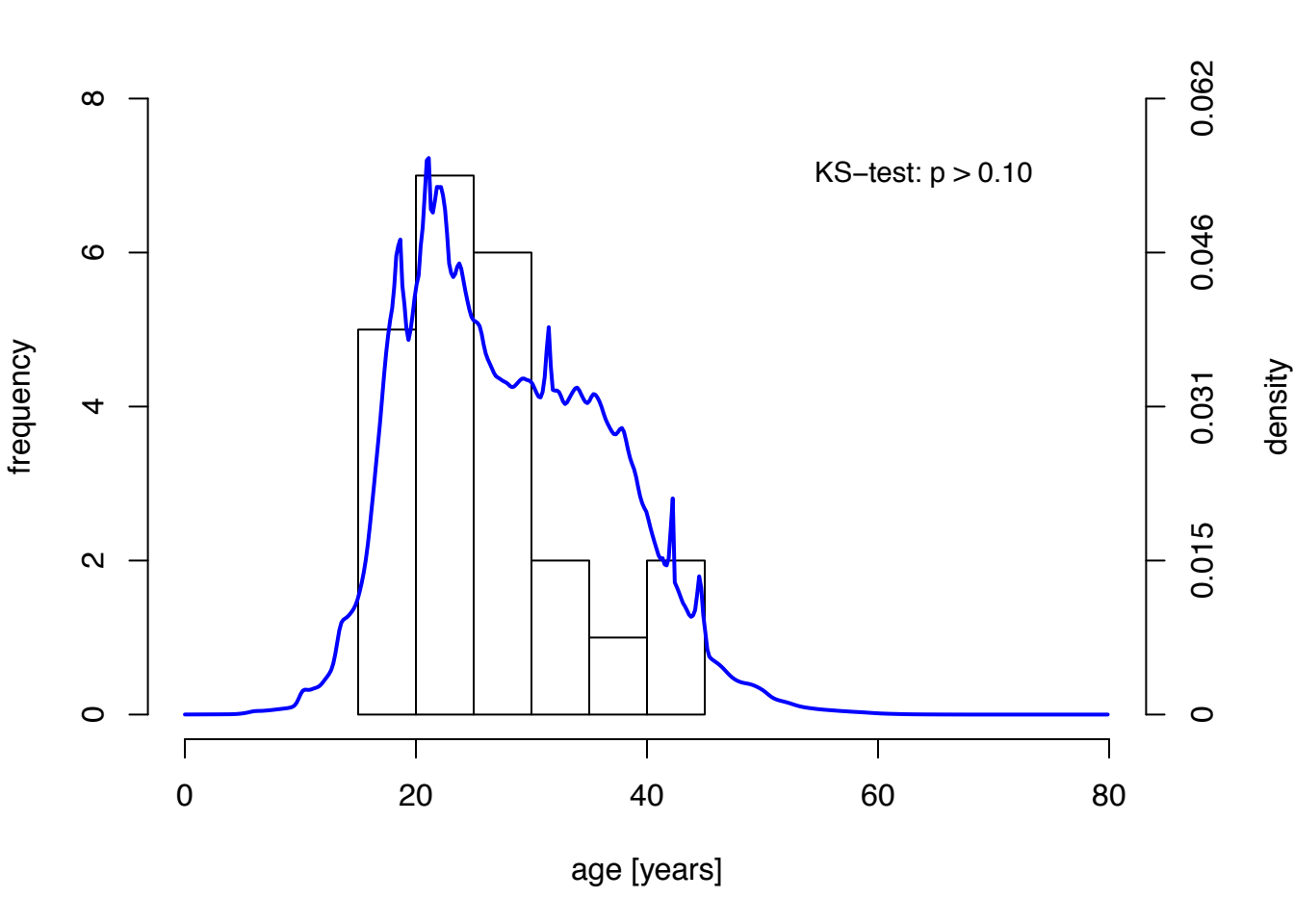
Figure 4. *Overall distribution of age at parturition for the Palanan Agta*. The overall distribution of the age at parturition in the Agta is obtained by averaging the age distributions obtained by the procedure depicted in Figure 5 for all pairs of mother and child in our Palanan Agta data set (blue density). This excludes 23 pairs for which the age of both mother and child are precisely known and that are shown separately (histogram).

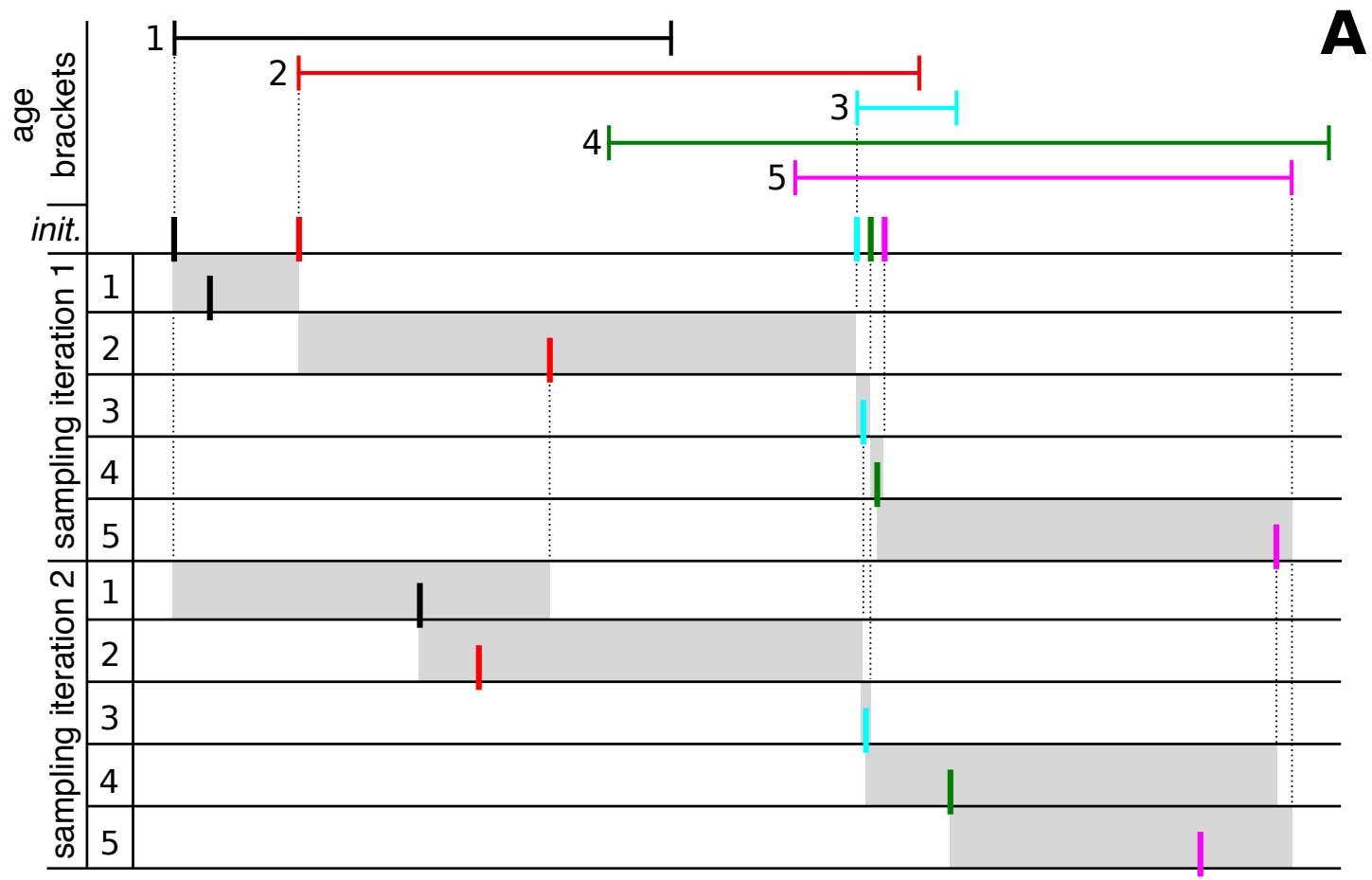
Figure 5. *Gibbs sampling of ages under ranking constraint (A) and exemplary input data and output of the Bayesian approach (B).* Panel A illustrates how the iterative sampling of ages works. Given age brackets, the age of each individual is initialised (init.) to the smallest possible value so that together the ages respect the ordering constraint (here 1 in black, 2 in red, 3 in cyan, 4 in green, and 5 in purple). Considering each individual in turn, a new age is sampled at random so that the ranking order remains valid at any time (admissible regions shaded in grey). See SI Appendix for the full mathematical description of the procedure. Panel B is a numerical example corresponding to individuals and age brackets from panel A and assuming uniform prior age distributions (in grey). The posterior distributions (same colour code as in panel A; here Kernel smoothed) generated by the iterative Gibbs sampling procedure described above are shown in black.

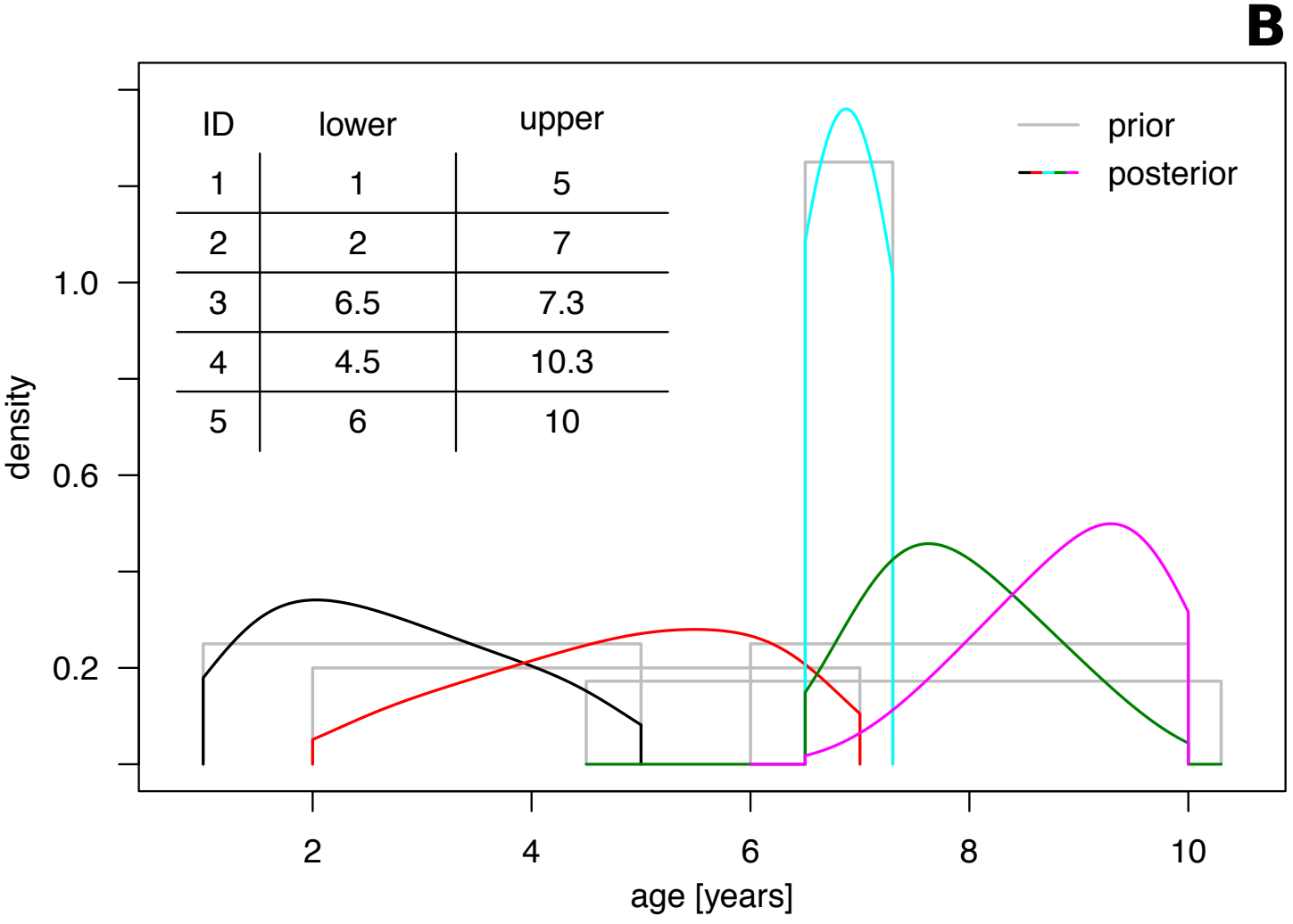

BF>419; p<1e−5
BF>2e12; p<1e−10
BF>4e20; p<1e−10
BF=0.23; p=0.61
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
1.16
4.39
0.64
0.29
0.33
mean
5th−order polynomial
5th−order polynnomial, with MP
LOESS
Bayesian (mean)
Bayesian (mean), no known ages


12444155554321434421111325235523221243334531514244134525513235213
known age
5th order polynomials
known age
age brackets with midpoint
5th order polynomials fitted incl. midpoints
known age
LOESS
known age
age brackets
Bayesian (means)
Bayesian (means), no known ages
age [years]
rank
0
20
40
60
80
1
10
20
30
40
50
60
65


child
mother
6
7
8
9
10
11
12
26
28
30
32
34
36
38
convolution
density
5
10
15
20
25
30
35
age [years]


KS−test: p > 0.10
frequency
density
age [years]
0
2
4
6
8
0.015
0.031
0.046
0.062
0
20
40
60
80


A
age brackets
1
2
3
4
5
init.
sampling iteration 1
sampling iteration 2


B
ID lower upper
1 1 5
2 2 7
3 6.5 7.3
4 4.5 10.3
5 6 10
prior
posterior
density
1.0
0.6
0.2
2
4
6
8
10
age [years]

SUPPLEMENTARY INFORMATION

Accurate age estimation in small-scale societies

Yoan Diekmann, Daniel Smith, Pascale Gerbault, Mark Dyble, Abigail E. Page, Nikhil Chaudhary, Andrea Bamberg Migliano, Mark G. Thomas

SUP. MATERIALS AND METHODS

Estimating ages by Gibbs sampling

We consider a random variable $\boldsymbol{X} = (X_1, \dots, X_n)$ with ages of n individuals. Furthermore, we introduce an ordering R of these n individuals from youngest to oldest, which can always be re-labeled such as $R = (1, \dots, n)$. In a Bayesian framework, age estimation can thus be formalized as computing the posterior distribution

$$P(\boldsymbol{X}|R) = \frac{P(R|\boldsymbol{X})P(\boldsymbol{X})}{\int_{X \in \boldsymbol{X}} P(R|\boldsymbol{X})P(\boldsymbol{X})d\boldsymbol{X}}$$

where $P(\boldsymbol{X})$ is an arbitrary prior distribution on the ages of the individuals satisfying $P(\boldsymbol{X}) = \prod_{i=1}^{n} P(X_i)$, and the likelihood function $P(R|\boldsymbol{X})$ is defined as

$$P\big(R = (1, \dots, n)\big|\boldsymbol{X} = (x_1, \dots, x_n)\big) = \begin{Bmatrix} 1 \; if \; x_i < x_j \; \forall \; i < j \\ 0 \; else \end{Bmatrix}.$$

In order to avoid explicit computation of the normalizing constant, we opted to approximate the posterior distribution by statistical sampling techniques. A naïve approach to sample from the posterior is to randomly draw an age for each of the n individuals independently, and then test if the resulting sample satisfies the ranking constraint. If not, the value is discarded. However, the more the individuals' prior age distributions overlap, the more samples generated by this approach would have to be discarded. To solve

this more efficiently, we implement a Gibbs sampling approach, which samples from the posterior distribution directly without having to discard any age-vector. The key to achieve this lies in considering only univariate conditional distributions, i.e. the age distribution of one individual when all other individuals are assigned a fixed value from their respective range (3, p. 16), i.e. $P_X(x_i \mid x_1, \dots x_{i-1}, x_{i+1}, \dots, x_n)$. How an initial set of values $\boldsymbol{x}$ satisfying the age ranking can be found is described below (point 1). Iterating over all individuals in this manner generates a sample $\boldsymbol{x}$, and it can be shown that the sequence of samples $\boldsymbol{x}$ thereby generated converges to the desired target posterior distribution (3, p. 17) .

In our case, a Gibbs sampler can be constructed in the following manner. First, we observe an ordering R of all individuals and label them accordingly, i.e. individual labeled 1 is younger than individual 2 etc., the oldest being individual n. Next, iterative rounds of sampling are performed. Denote the k^{th} sample of ages $\boldsymbol{x}$ by $\boldsymbol{x}^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$. Assume for example that $P_X(x) \sim Unif(l, u)$, i.e. the *a priori* age of any individual is distributed uniformly within an interval bounded by values l and u. We note that alternative distributions for $P_X(x)$ – such as a normally distributed *a priori* age – are easily accommodated in a way analogous to the one described below. Setting $x_0 \coloneqq -\infty$ and $x_{n+1} \coloneqq \infty$ for the sake of simplicity, our Gibbs sampler proceeds as follows:

1) Initialize the first sample $k = 0$:

$x_i^{(0)} = \max(l_i, x_{i-1}^{(0)})$, for $i \in \{1, \dots, n\}$

2) Iterate K times to generate $K + 1$ samples, i.e. $k \in \{1, \dots, K\}$:

$x_i^{(k)} \sim Unif\left(\max\left(l_i, x_{i-1}^{k}\right), \min\left(u_i, \dots, u_n, x_{i+1}^{(k-1)}\right)\right)$, for $i \in \{1, \dots, n\}$

This procedure generates as many samples as desired. As always with empirical distributions, the general trade-off is that more samples occupy more memory space and require longer computation time, but reduce the stochastic sampling error and therefore better approximate the underlying distribution.

Figure 5 in the main text illustrates the type of input required and output generated by our method for five fictitious individuals.

Implementation details

We have implemented the Gibbs sampling algorithm in Python 2.7 (5). In order to find sensible parameter values for the total number of iterations, burn-in and thinning, we analysed 50,000 sampling iterations for the toy example with five individuals presented in Figure 5B of the main text.

Panel A of Supplementary Figure S3 shows perfect mixing, with low autocorrelation (see Panel D) also confirmed by a high effective sample size of 33521.62, meaning that for the estimation of the posterior mean 50,000 samples correspond to 33,522 independent samples. This suggests that no thinning is required. Panel B and C illustrate how the sample mean changes in the course of the sampling process. Based on visual inspection, we chose a burn-in of 50 iterations, largely exceeding Raftery-Lewis (9) method's recommendation of two to four. Panel B already suggests that convergence is achieved relatively quickly, as means remain stable after 10,000 iterations. Gelman and Rubin's shrink factor (8), a formal test for convergence presented in Panels E and F and computed on 4 independent runs of the Gibbs sampler with the first 10,000 iterations discarded, shows a shrink factor of 1 after 10,000 additional iterations. Therefore, we set our default to 20,050 iterations in total, resulting in 20,000 ages sampled per individual with no thinning and 50 iterations discarded a burn-in.

All diagnostic statistics were computes and plotted in R version 3.1.3 (6) using functions from the 'coda' and 'mcmcplots' libraries.

Palanan Agta: data collection method

In order to construct relative age rankings, we took and printed photographs of all individuals in every camp. Individuals were then assigned to approximate age cohorts (0-4, 4-8, 8-12, 13-19, 20-45, and 45+). Those not easily assigned to one cohort were included in the two nearest cohorts (e.g.,

an individual aged ~45 would be included in both the 20-45 and 45+ cohorts). Either individually or in small groups, we presented these photographs to individuals from a target cohort, one at a time. The target cohort was the cohort the individual ('ego') was included in, as well as all cohorts younger than ego. Cohorts, especially for children, were often presented together, so that some rankings included, for instance, all individuals aged 0 to 12. Children under the age of five were often unable to make the age rankings themselves, and in this instance either their mothers or older siblings would conduct the ranking. Individuals from a specific camp were shown pictures of others from their camp and neighbouring camps. More distant camps were not included due to a lack of familiarity, unless ego knew individuals from more distant camps particularly well (e.g. they grew up in the same camp and moved apart upon marriage). For cohorts including ego, ego's picture was displayed first. Participants were first asked if they knew the individual on the photograph (i.e. the target), and if so they were then asked if they knew the target well enough to give their approximate date of birth relative to other individuals. Each photograph was put into one of three categories; 'don't know', 'know but not the age', and 'age known'. If ego knew both the target and their age, they were asked to rank the age of the target relative to others. Although similar to the method by Hill and Hurtado (2), rather than having two piles of simply older and younger (with ego as reference), our method produced a relative age list from youngest to oldest. This process was repeated multiple times with different subjects producing a total of 266 partial ranks, including 587 individuals.

The second stage involved deriving age estimates for these 587 individuals. One invaluable source of information, especially for older individuals, was the Headlands' database from Casiguran (4), since some individuals from our study population were included in this database, with relatively accurate dates of birth assigned. Absolute ages of individuals were ascertained via various other methods, including; asking individuals if they knew their own or their children's age (which could be from various sources, such as, birth certificates, other documentation, school grades, own estimates, etc.), births near dated events (such as martial law in 1970 or

various known typhoons), and age-mates of individuals with known birthdays. For children up to the age of 12 years, it was also possible to estimate age brackets by dental development.

There are, however, some issues with methods used to estimate absolute ages, especially estimates given by individual Agta, the dental aging and school grade. For example, many individuals gave various conflicting dates and/or ages, including; saying a child was four years old, yet born in 2004 (during the 2013 fieldwork season), or giving a birth date for one child as 2004 (~eight years old) yet saying a younger child was nine years old, and age conflicts between parents (for example, one child was given an age of seven months by one parent and two years by the other). For both teeth ages and school grades, the margins of error were often quite large (+/- half a year), which was especially problematic regarding school ages, as the grade reached was often variable for individuals of a similar age, and most children in the community either do not go to school, or start school at older ages than their agricultural neighbours. Therefore, strict criteria were used to select accurate ages/birth dates. First, if an individual was given two markedly different birth dates, that person was excluded from the absolute age list. Second, if ages for an entire sibling-set were provided, but at least one age was wrong (e.g., did not correspond to teeth ages, or did not allow at least nine months pre- or post-birth of the nearest sibling), then ages for the whole sibling-set were excluded. Furthermore, for all children, the birth date had to fall within the range of teeth ages to be accepted, and a similar protocol of matching with teeth ages was established for estimating the ages of individuals from school grade. For ages estimated based on comparisons to individuals with known birth dates, these individuals with estimated ages were given a year of birth with a +/- one year margin to account for error. Using these methods, 98 individuals (out of 587; 16.7%) were given an exact birthday, while many others were given age estimates within +/- one year (Supplementary Table S3).

For individuals which we could not attach a secure date or estimate, three of the field researchers (DS, AEP, & MD), as well as the principle investigator (ABM) estimated the ages based on cues such as dental

development, school grade, birth order (if older or younger siblings have a known age), age of ego's children (if known), number of children, and visual inspection. Independently, each of the four researchers estimated an upper and lower age bound for each individual. In collating these estimates, the youngest lower bound and oldest upper bound of the four estimates were used in order to include as much uncertainty as possible. There was increased uncertainty for older individuals, as the average difference between upper and lower estimates increases with age (Supplementary Table S3).

SUP. RESULTS

Validation and benchmarking

Table 1 and Figure 1 in the main text show that the Gibbs sampler provides more accurate age estimates than the regression approach. However, the performances may be influenced by the specific cross-validation parameters chosen, i.e. k=5 partitions of n=13 individuals each for which ages are assumed to be known exactly. Therefore, we tested other parameter values from k=2 partitions, resulting in n=32 individuals, to k=13, with n=5 individuals per partition. We considered each partition in turn to estimate the regression equation and then deduced the ages of the remaining individuals. This procedure enabled us to assess how the number of individuals with known ages affects each method's accuracy.

Supplementary Figure S1 shows that the accuracy for the fifth-degree polynomial approach massively drops when more than five partitions are chosen (i.e. k>5). This is expected, as fewer known ages are available for the regression, resulting in a less constraint curve leading to overfitting. Note that although the LOESS approach also shows reduced accuracy in smaller partitions, the magnitude of the error is much smaller.

A flexible method for fieldwork data: dealing with multiple partial ranks

We relax the assumption of a single complete ordering R of all n individuals from youngest to oldest, and rather allow for multiple partial ranks. The approach we describe in the following is heuristic. Describing the problem of multiple partial ranks in a formal manner and finding optimal solutions is an important and interesting problem for future research.

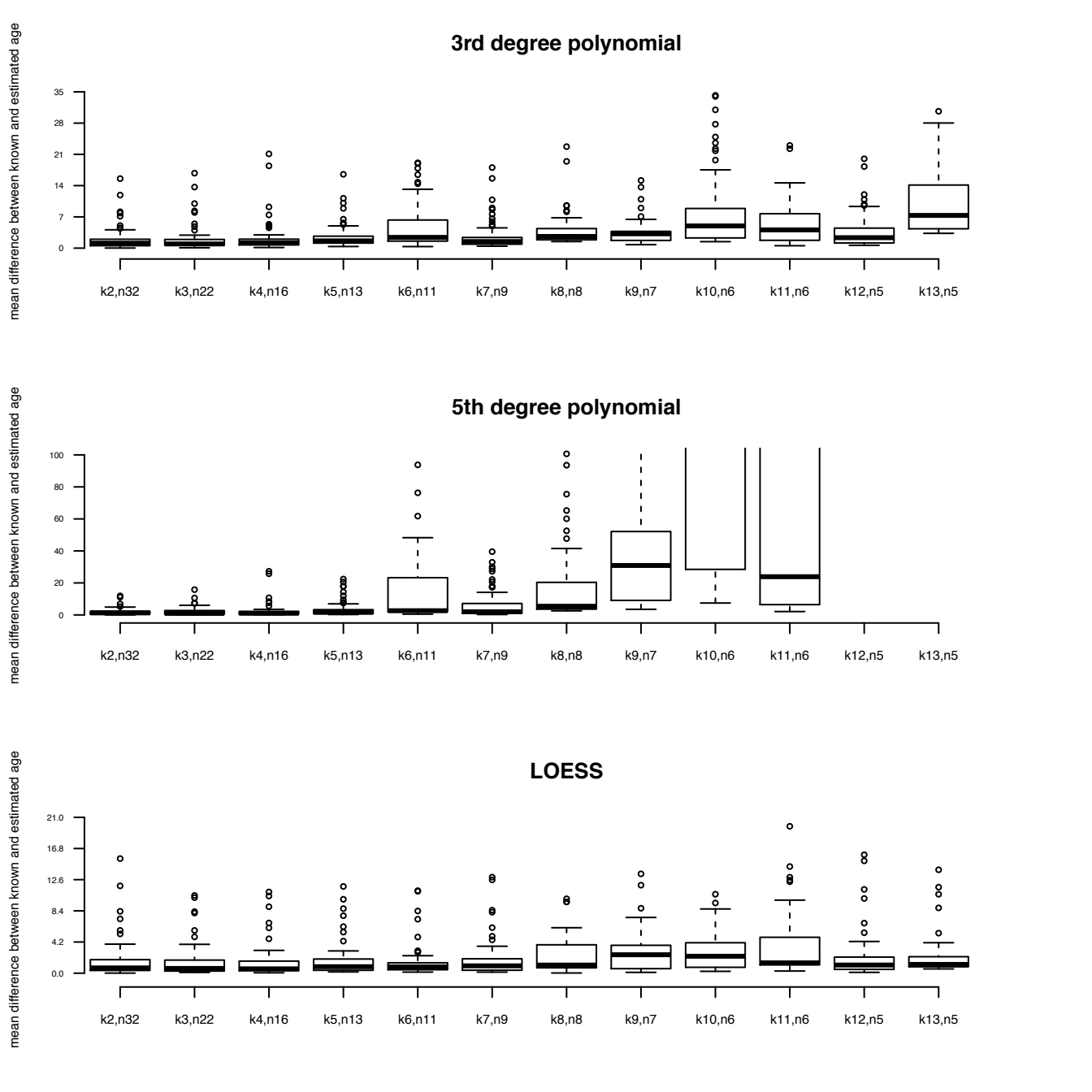
Let $\mathfrak{R} = \{R_1, \dots, R_m\}$ be a set of partial rankings of individuals. As described in the main text, we first merge partial ranks that are compatible, resulting in a modified set of partial ranks $\{\boldsymbol{R'}_1, \dots, \boldsymbol{R'}_l\}$, $l \leq m$, where each $\boldsymbol{R'}_j$ represents a subset of mutually compatible partial ranks from the initial full set, i.e. $\boldsymbol{R'}_j \subseteq \mathfrak{R}$**.** Merging is not always possible without ambiguity, as various different ways in which rankings could be merged may exist, e.g. if R_1 is compatible with R_2 and R_3, but R_2 and R_3 are not compatible with each other. In this case, we leave the corresponding ranks separate ($\{\boldsymbol{R'}_1, \dots, \boldsymbol{R'}_l\}$ is therefore a partition of the set $\mathfrak{R}$). It should be noted that alternative heuristics can easily be envisaged at this stage, for example a greedy strategy. The next step is to compute the posterior $P\left(\boldsymbol{X}\middle|\boldsymbol{R'}_j\right)$ separately for all merged partial ranks $\boldsymbol{R'}_j, j \in \{1, \dots, l\}$, by Gibbs sampling. Finally, we merge the resulting distributions per individual by forming a weighted finite mixture:

$$P(X_i = x|\mathfrak{R}) = \sum_{j=1}^{l} \frac{w_i(\boldsymbol{R'}_j)}{w_i(\mathfrak{R})} P\left(X_{\boldsymbol{i}} = x\middle|\boldsymbol{R'}_{\boldsymbol{j}}\right)$$

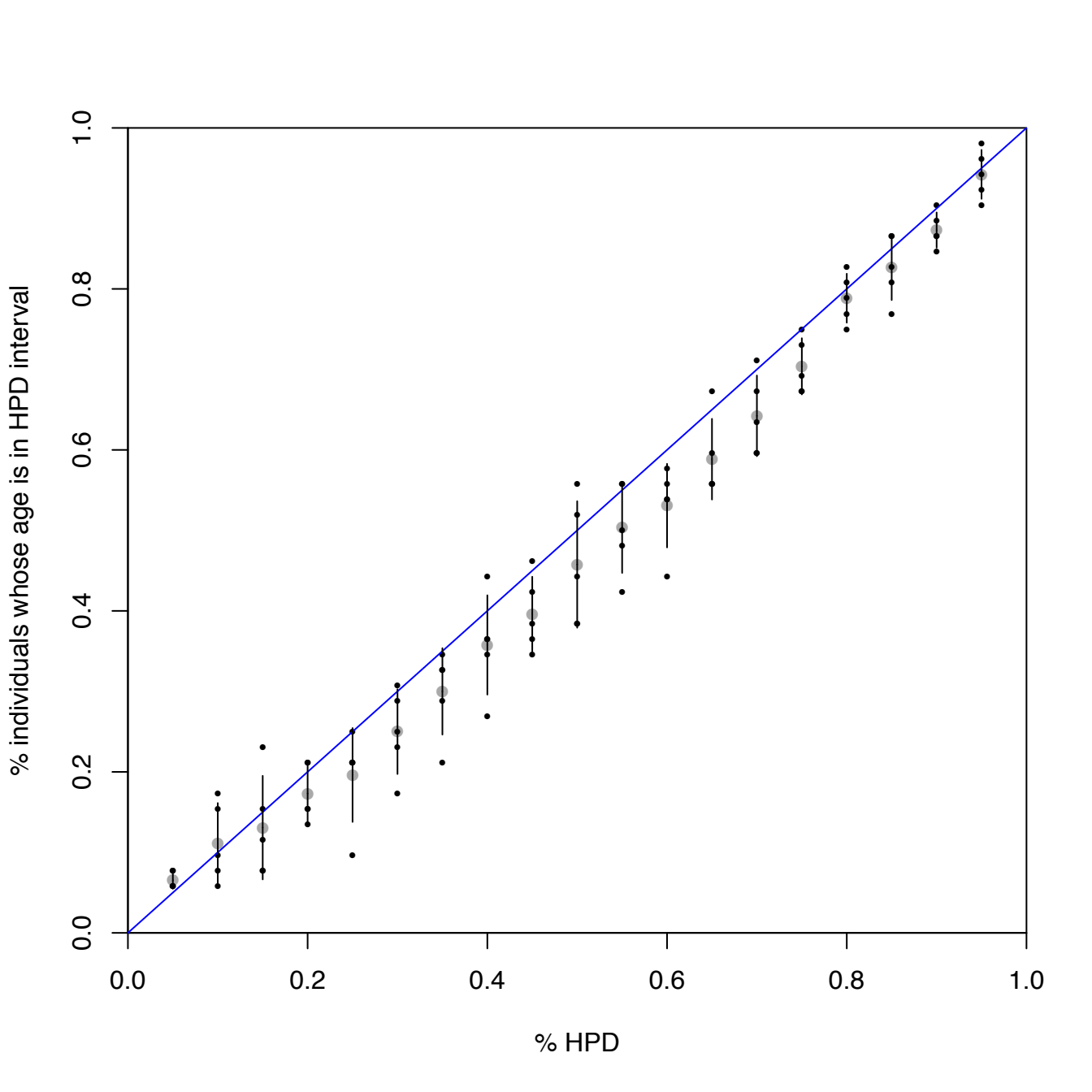
where $w_i()$ denotes the number of times individual i occurs in the corresponding set of rankings. The nominator term $w_i(\boldsymbol{R'})$ therefore preserves the information how many times an individual has been ranked consistently in a certain way in the initial set of unmerged partial rankings $\mathfrak{R}$.

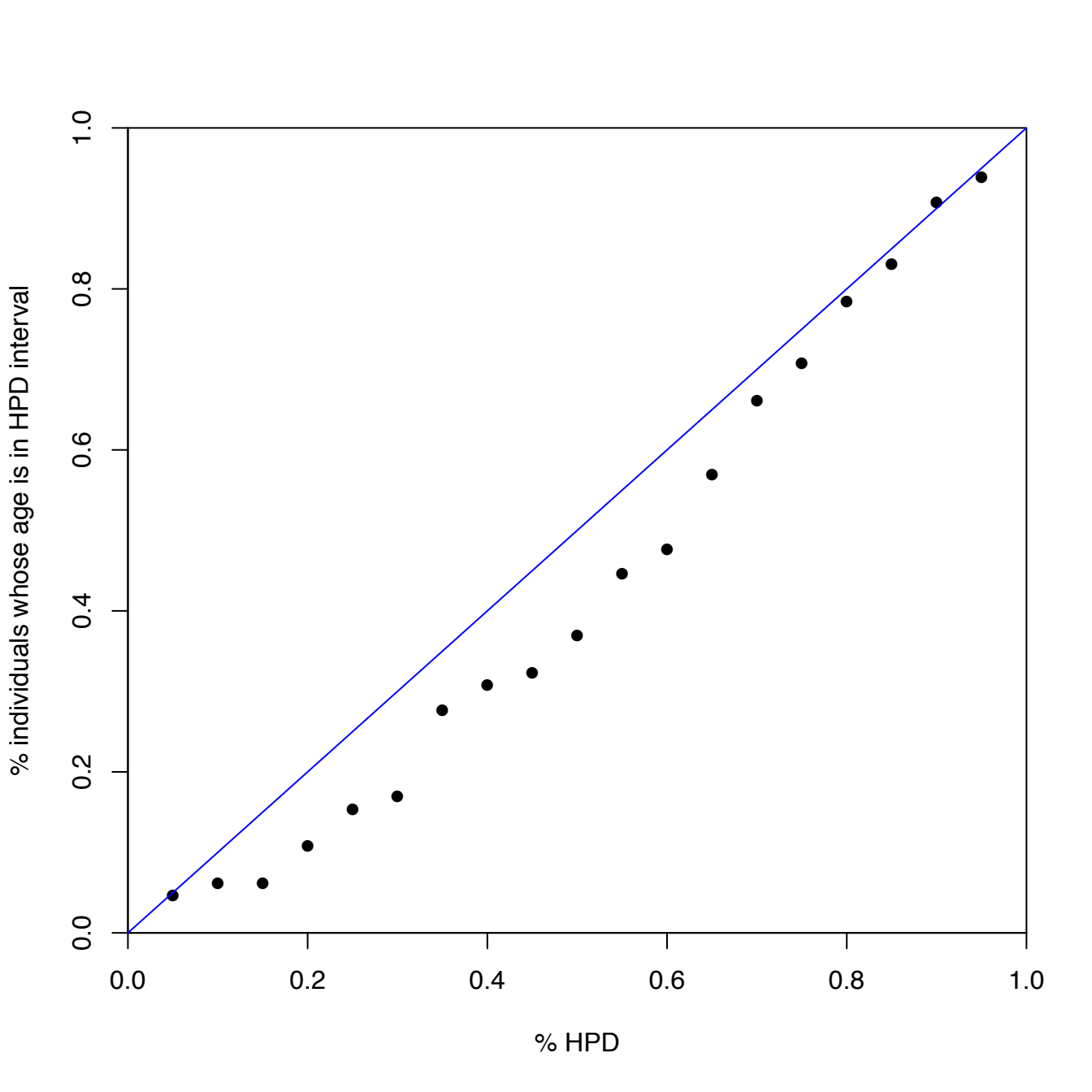
SUP. FIGURES

Supplementary Figure S1. *Differences in estimation accuracy under varying cross-validation parameters.* Boxplots of the mean of the differences between known ages and those estimated using regression analyses; top: third-order (3rd degree) polynomial, middle: fifth-order (5th degree) polynomial, bottom: local regression (LOESS; 7). The x-axis shows the number of partitions used ('k') and the number of individuals ('n') in these corresponding partitions; 'k2,n32' for example means 2 partitions of 32 individuals whose ages are known and used to estimate the regression coefficients. The y-axis shows the mean of the differences between known and estimated ages per individuals over the k partitions. Note that the scale of the y-axis of these three panels is not the same.

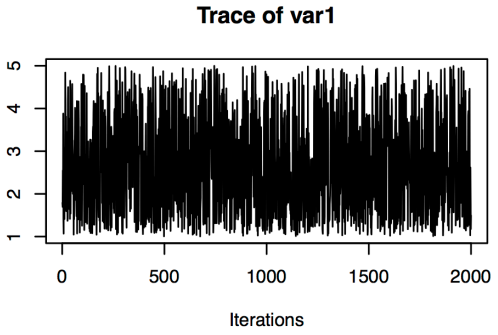

3rd degree polynomial
mean difference between known and estimated age
0
7
14
21
28
35
k2,n32
k3,n22
k4,n16
k5,n13
k6,n11
k7,n9
k8,n8
k9,n7
k10,n6
k11,n6
k12,n5
k13,n5
5th degree polynomial
mean difference between known and estimated age
0
20
40
60
80
100
k2,n32
k3,n22
k4,n16
k5,n13
k6,n11
k7,n9
k8,n8
k9,n7
k10,n6
k11,n6
k12,n5
k13,n5
LOESS
mean difference between known and estimated age
0.0
4.2
8.4
12.6
16.8
21.0
k2,n32
k3,n22
k4,n16
k5,n13
k6,n11
k7,n9
k8,n8
k9,n7
k10,n6
k11,n6
k12,n5
k13,n5

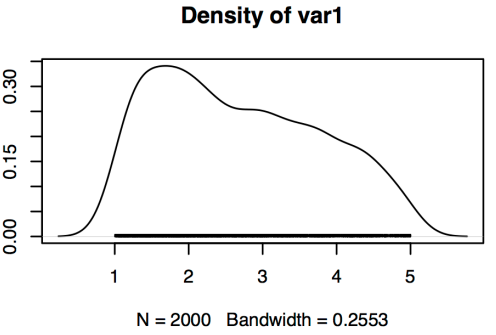
Supplementary Figure S2. *Error calibration of posterior distributions.* For the cross-validation experiment corresponding to Figure 1, we show that the highest posterior densities (HPD) contain the true age as often as the size of the interval suggests, and the posterior therefore correctly quantifies estimation uncertainty. For example, the 95% HPD covers the true age in 95% of the individuals. Panel A shows the results for each of the 5 cross-validation partitions (black points), their average (grey points) and standard deviation (black bars). Panel B shows the same analysis for the case where no age has been fixed, i.e. all priors were proper intervals.

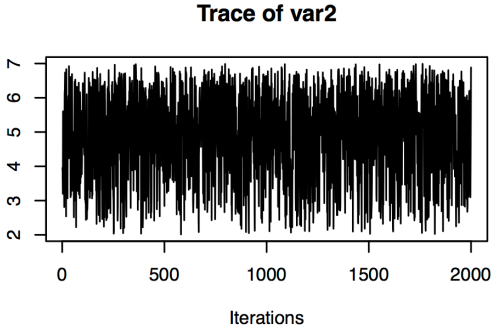

% individuals whose age is in HPD interval
0.0
0.2
0.4
0.6
0.8
1.0
% HPD
0.0
0.2
0.4
0.6
0.8
1.0

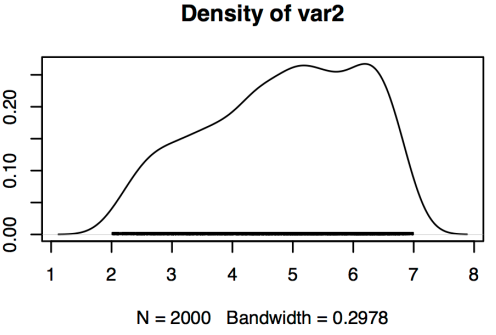

% individuals whose age is in HPD interval
0.0
0.2
0.4
0.6
0.8
1.0
% HPD
0.0
0.2
0.4
0.6
0.8
1.0

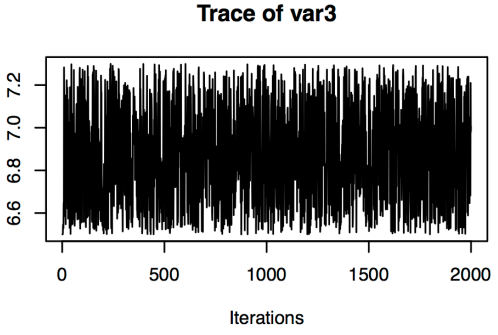
Supplementary Figure S3. *Gibbs sampler diagnostic statistics.* 50,000 sampling iterations were performed for the toy example with five individuals presented in Figure 5B of the main text. In Panels A to D, all sampling iterations are included, i.e. no burn-in is discarded. Panel A shows the trace and resulting density estimates (less smoothed versions of densities shown in Figure 5B) for the first 2000 iterations. Panel B and C show the running mean age for all 50,000 respectively for the first 500 samples. Panel D visualises the autocorrelation between consecutive samples. Panel E and F show Gelman and Rubin's shrink factor (8) on all respectively the first 2000 samples after discarding the first 10,000 in 4 independent runs of the Gibbs sampler.

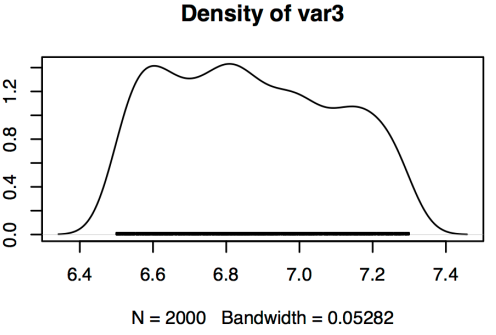

Trace of var1
1
2
3
4
5
0
500
1000
1500
2000
Iterations

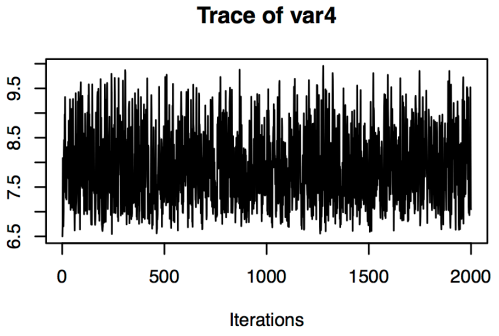

Density of var1
0.00
0.15
0.30
1
2
3
4
5
N = 2000 Bandwidth = 0.2553

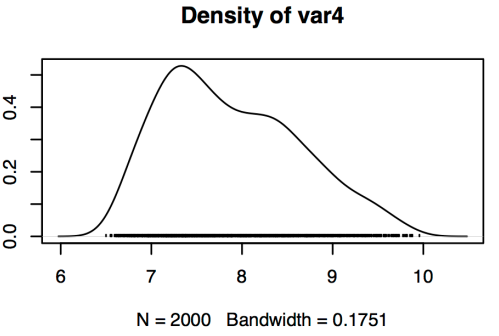

Trace of var2
2
3
4
5
6
7
0
500
1000
1500
2000
Iterations

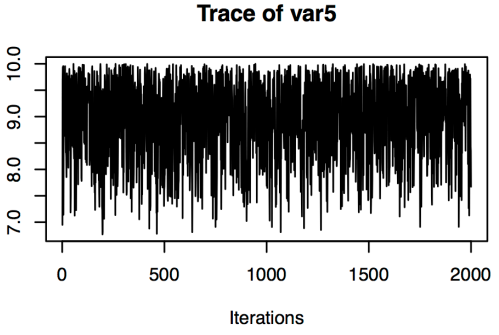

Density of var2
0.00
0.10
0.20
1
2
3
4
5
6
7
8
N = 2000 Bandwidth = 0.2978

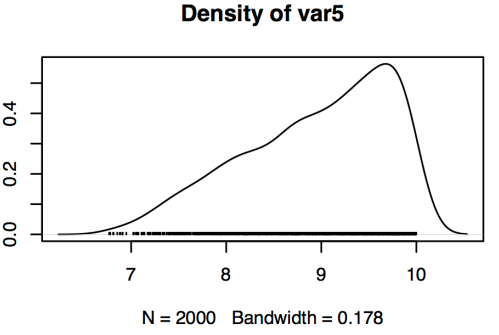

Trace of var3
6.6
6.8
7.0
7.2
0
500
1000
1500
2000
Iterations

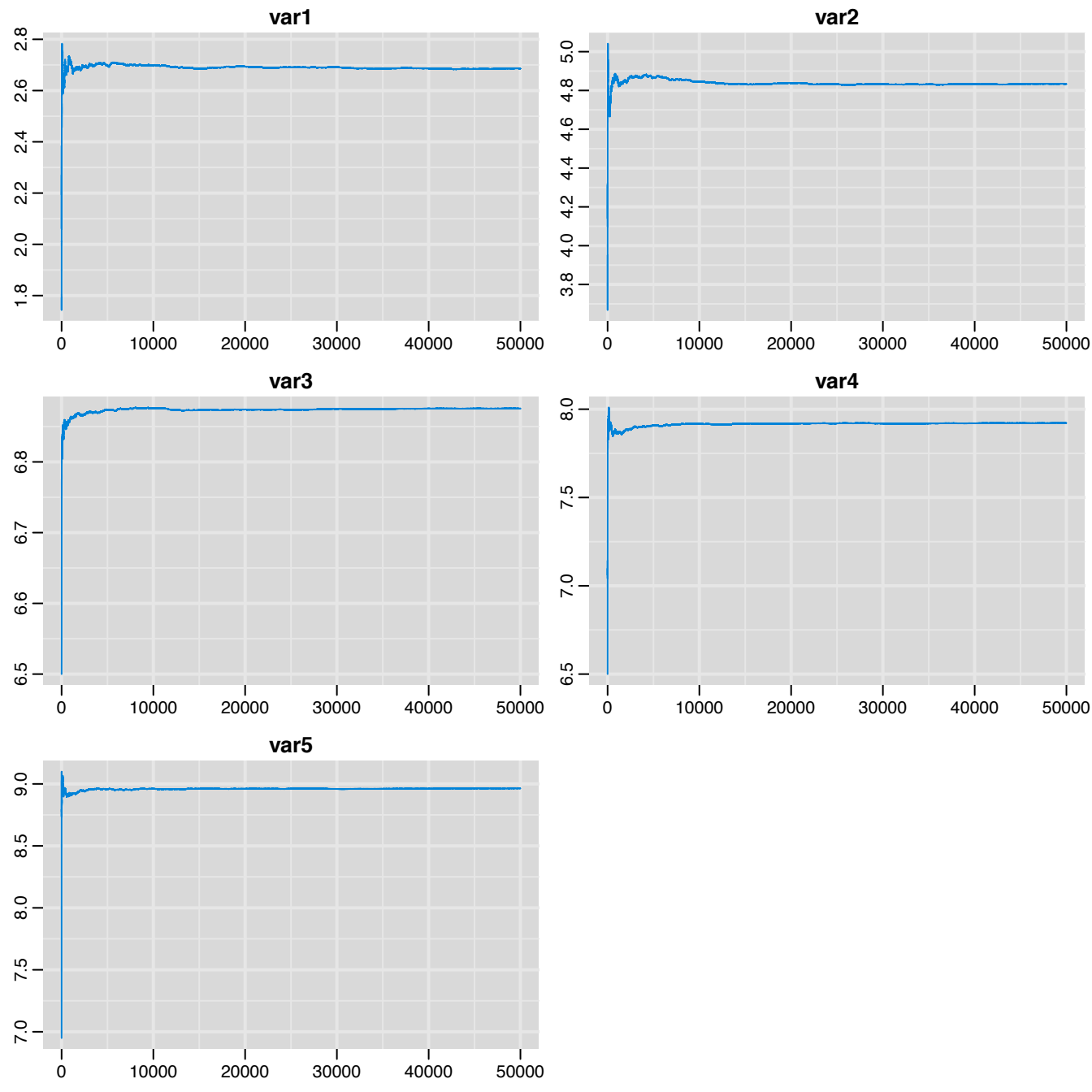

Density of var3
0.0
0.4
0.8
1.2
6.4
6.6
6.8
7.0
7.2
7.4
N = 2000 Bandwidth = 0.05282

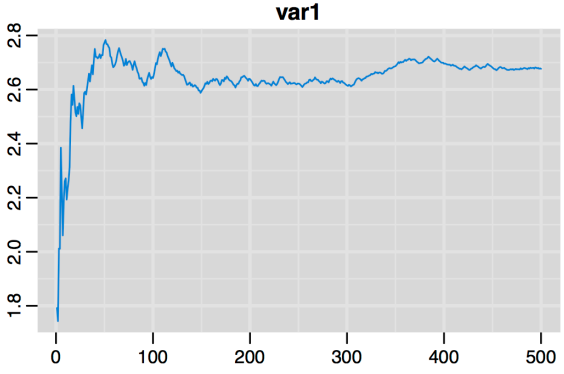

Trace of var4
6.5
7.5
8.5
9.5
0
500
1000
1500
2000
Iterations

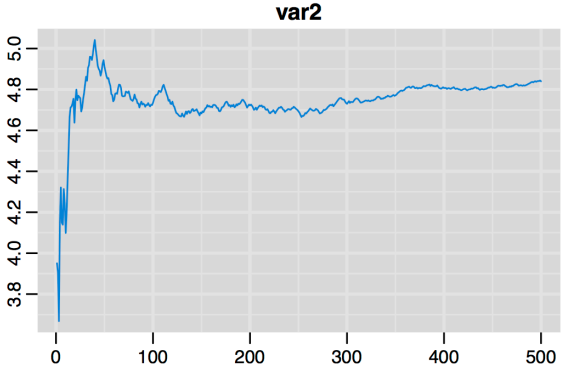

Density of var4
0.0
0.2
0.4
6
7
8
9
10
N = 2000 Bandwidth = 0.1751

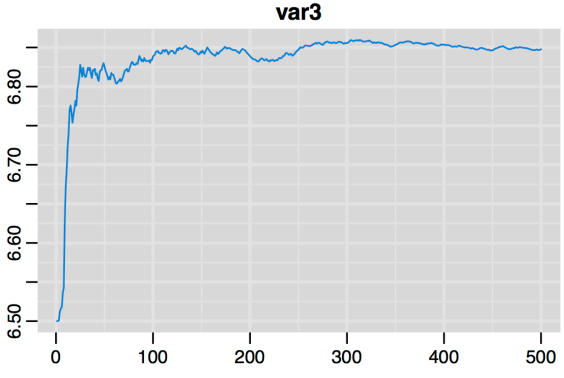

Trace of var5
7.0
8.0
9.0
10.0
0
500
1000
1500
2000
Iterations

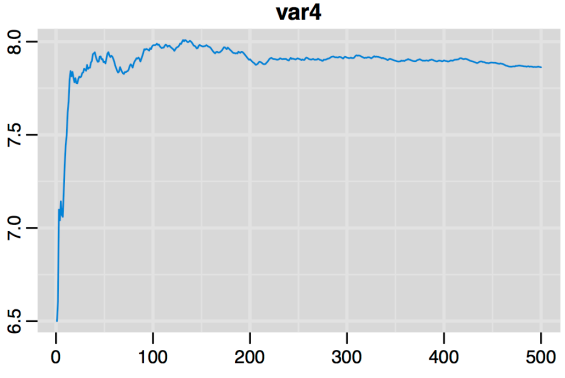

Density of var5
0.0
0.2
0.4
7
8
9
10
N = 2000 Bandwidth = 0.178

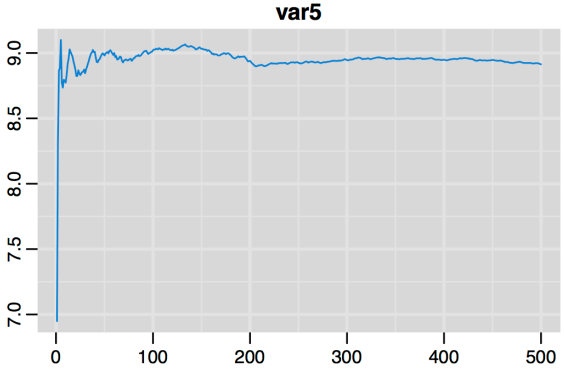

var1
1.8
2.0
2.2
2.4
2.6
2.8
0
10000
20000
30000
40000
50000
var2
3.8
4.0
4.2
4.4
4.6
4.8
5.0
0
10000
20000
30000
40000
50000
var3
6.5
6.6
6.7
6.8
0
10000
20000
30000
40000
50000
var4
6.5
7.0
7.5
8.0
0
10000
20000
30000
40000
50000
var5
7.0
7.5
8.0
8.5
9.0
0
10000
20000
30000
40000
50000

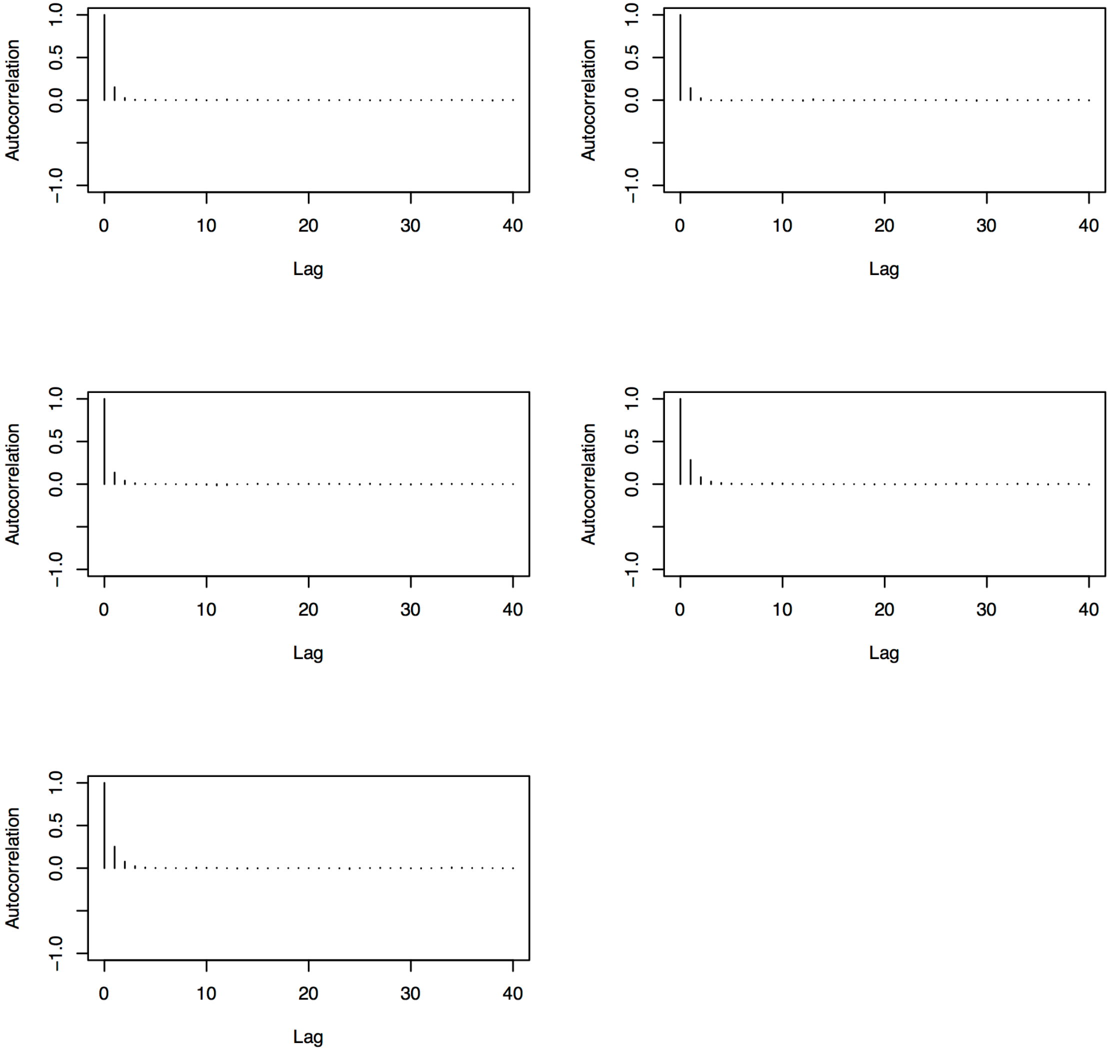

var1
1.8
2.0
2.2
2.4
2.6
2.8
0
100
200
300
400
500

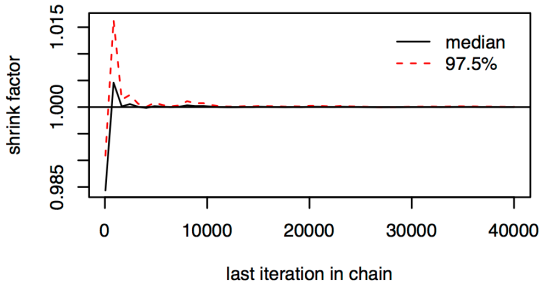

var2
3.8
4.0
4.2
4.4
4.6
4.8
5.0
0
100
200
300
400
500

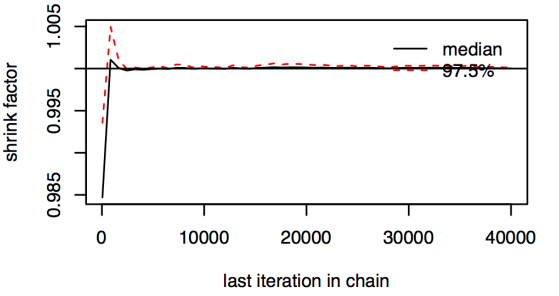

var3
6.50
6.60
6.70
6.80
0
100
200
300
400
500

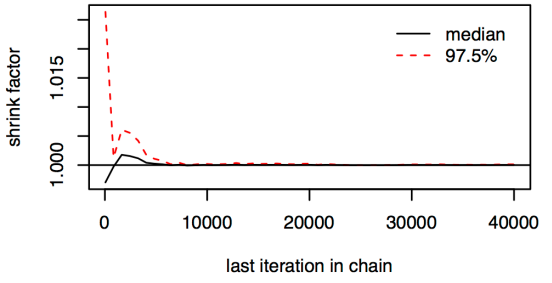

var4
6.5
7.0
7.5
8.0
0
100
200
300
400
500

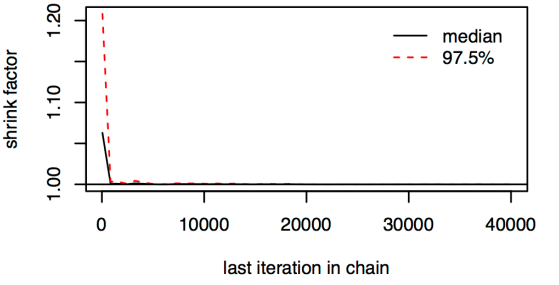

var5
7.0
7.5
8.0
8.5
9.0
0
100
200
300
400
500

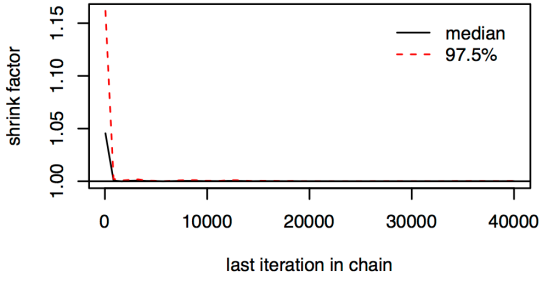

Autocorrelation
1.0
0.5
0.0
−1.0
0
10
20
30
40
Lag

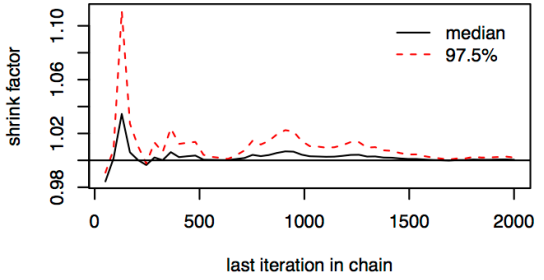

shrink factor
0.985
1.000
1.015
median
97.5%
0
10000
20000
30000
40000
last iteration in chain

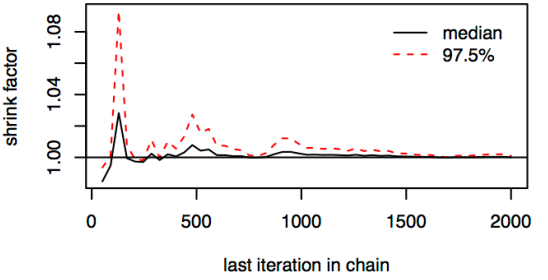

shrink factor
0.985
0.995
1.005
median
97.5%
0
10000
20000
30000
40000
last iteration in chain

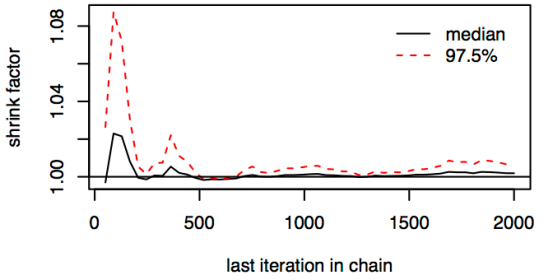

shrink factor
1.000
1.015
median
97.5%
0
10000
20000
30000
40000
last iteration in chain

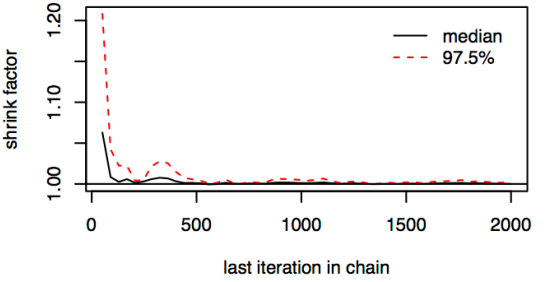

shrink factor
1.00
1.10
1.20
median
97.5%
0
10000
20000
30000
40000
last iteration in chain

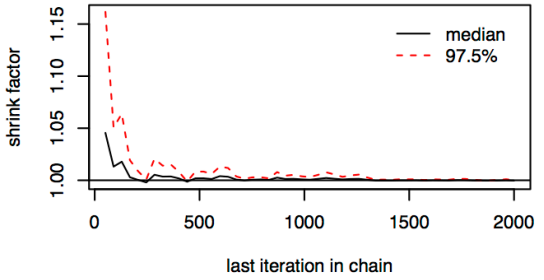

shrink factor
1.00
1.05
1.10
1.15
median
97.5%
0
10000
20000
30000
40000
last iteration in chain


shrink factor
0.98
1.02
1.06
1.10
median
97.5%
0
500
1000
1500
2000
last iteration in chain

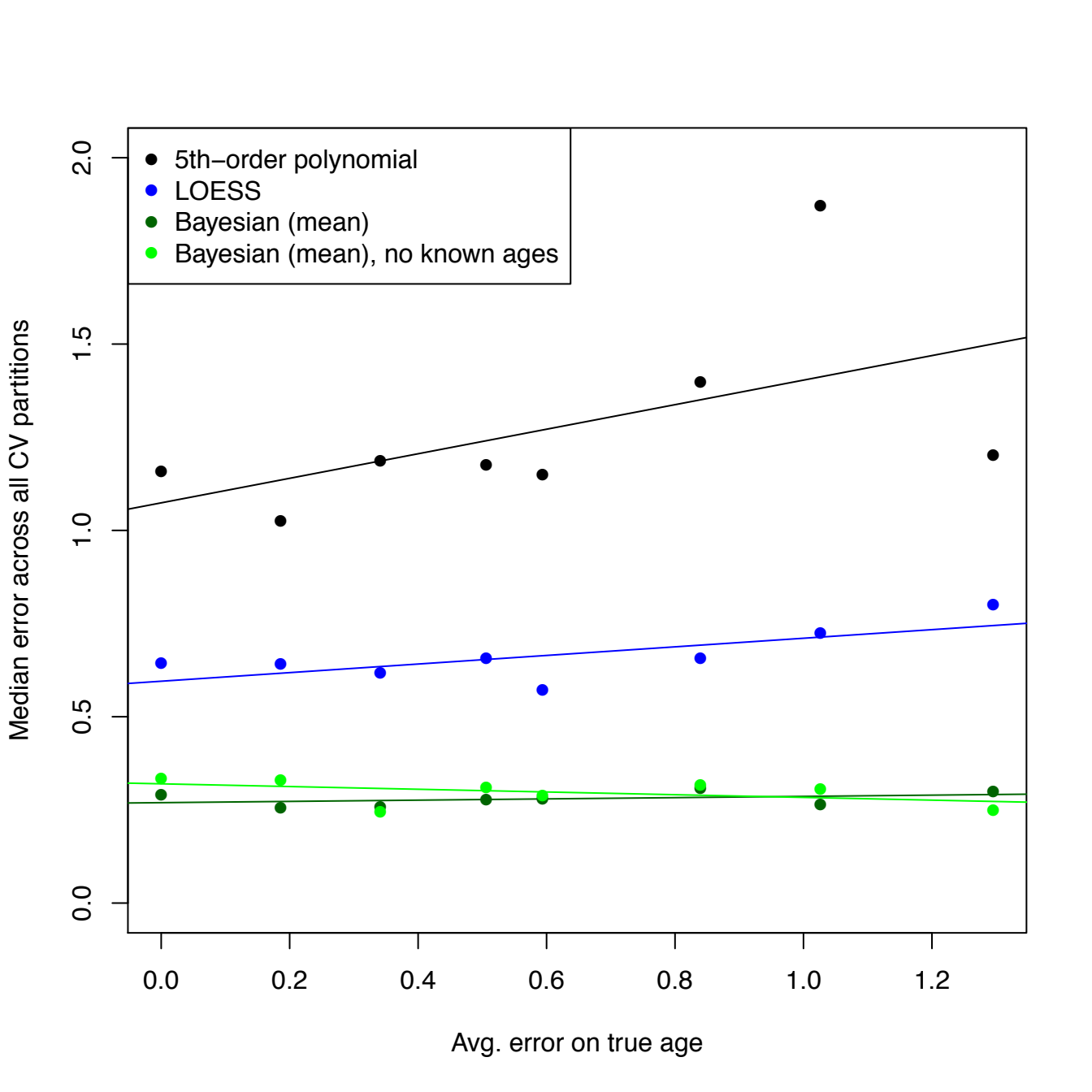

shrink factor
1.00
1.04
1.08
median
97.5%
0
500
1000
1500
2000
last iteration in chain

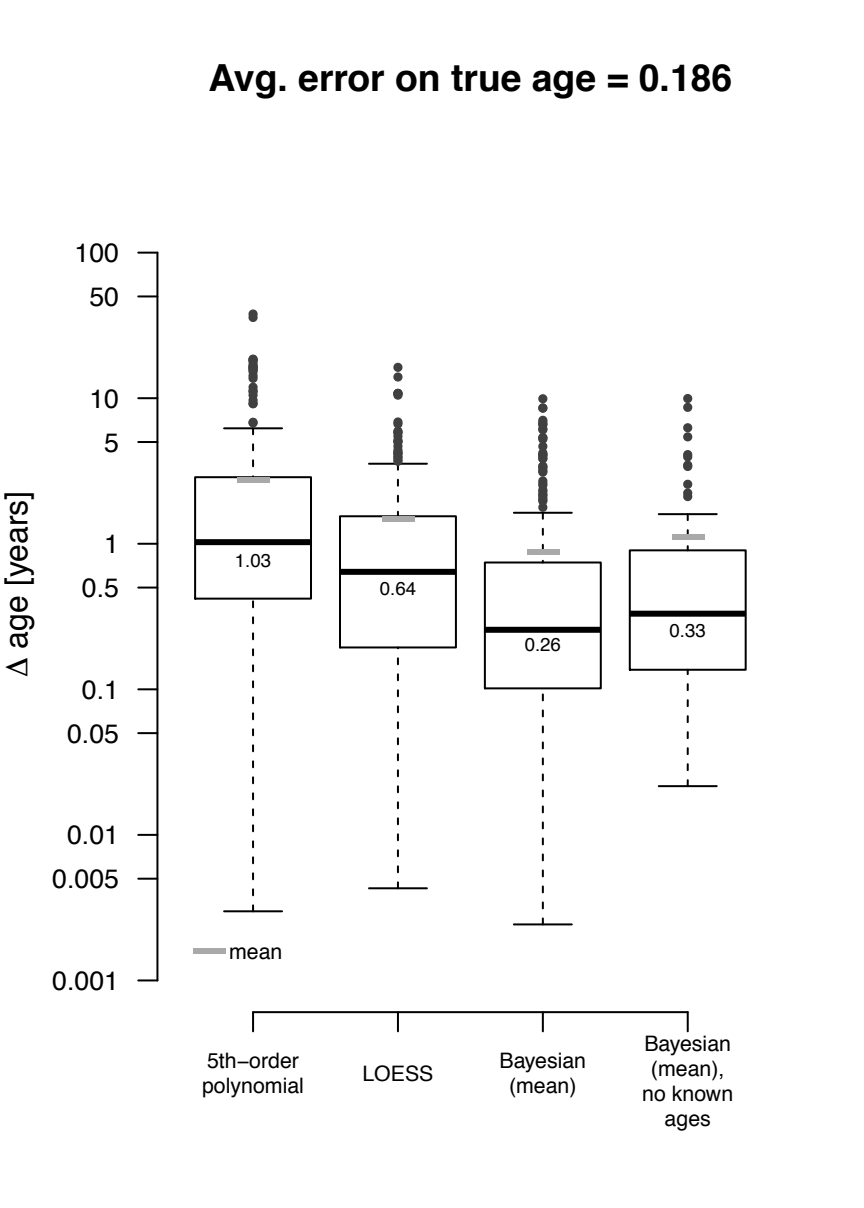

shrink factor
1.00
1.04
1.08
median
97.5%
0
500
1000
1500
2000
last iteration in chain

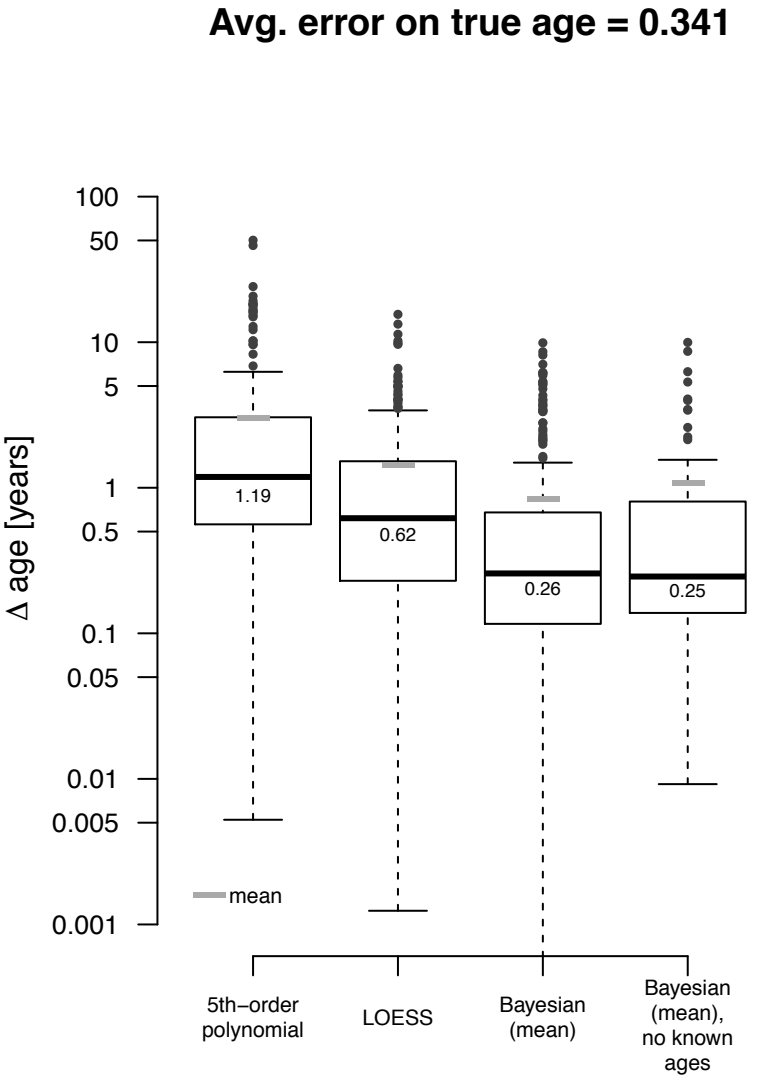

shrink factor
1.00
1.10
1.20
median
97.5%
0
500
1000
1500
2000
last iteration in chain


shrink factor
1.00
1.05
1.10
1.15
median
97.5%
0
500
1000
1500
2000
last iteration in chain

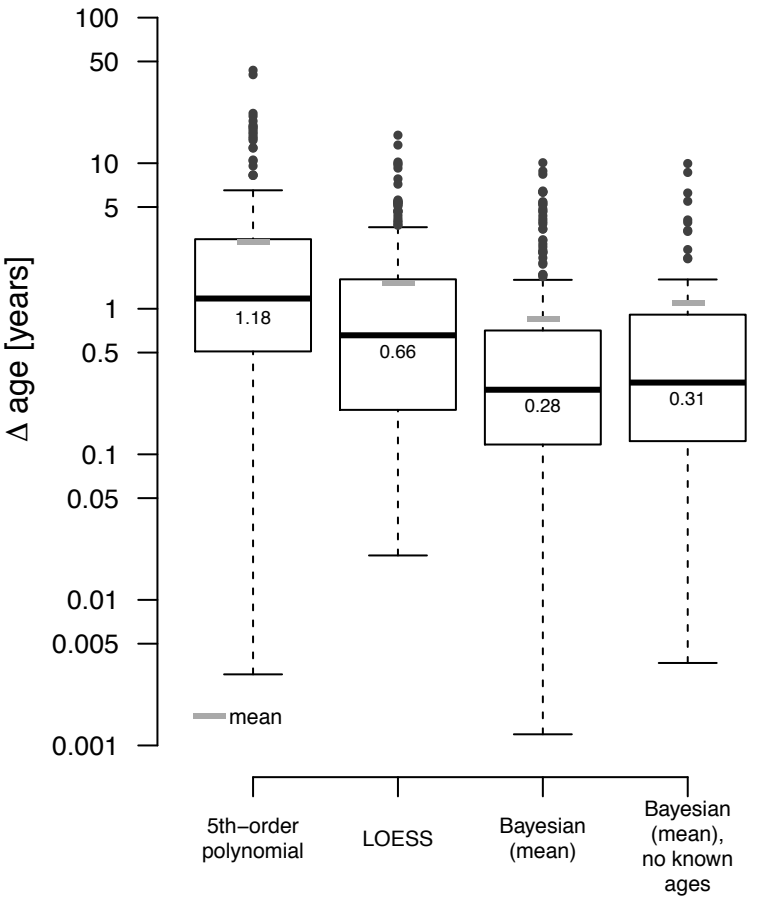
Supplementary Figure S4. *Estimation robustness to error in known ages.* We repeated the validation from Figures 1 and 2, however, added different amounts of error to the individuals' ages, where errors are constrained not to change the ranking order. Panel A summarizes how this affects the different methods: linear regression shows that estimation accuracy measured as the median of the differences between estimated and actual ages of the individuals across the 5 cross-validation partitions is reduced most for the polynomial regression approach, slightly for LOESS and not at all for our Bayesian method. Panel B gives the corresponding distributions in form of boxplots.

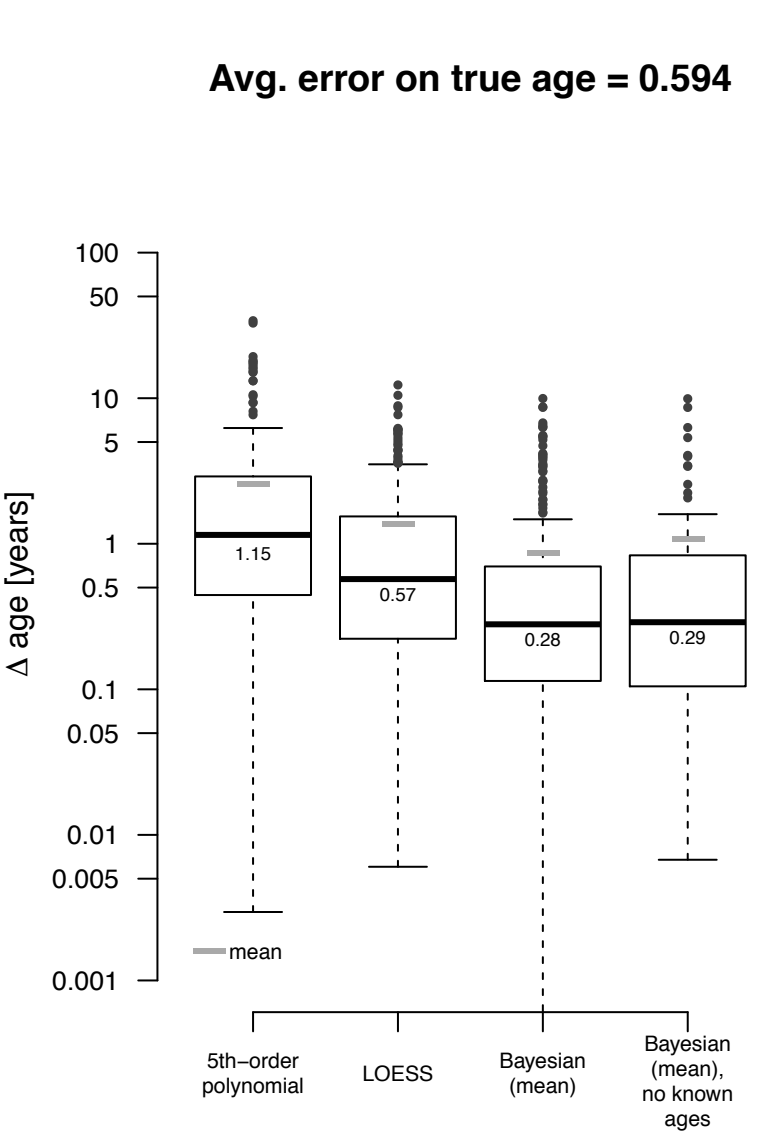

5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages
Median error across all CV partitions
Avg. error on true age
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2

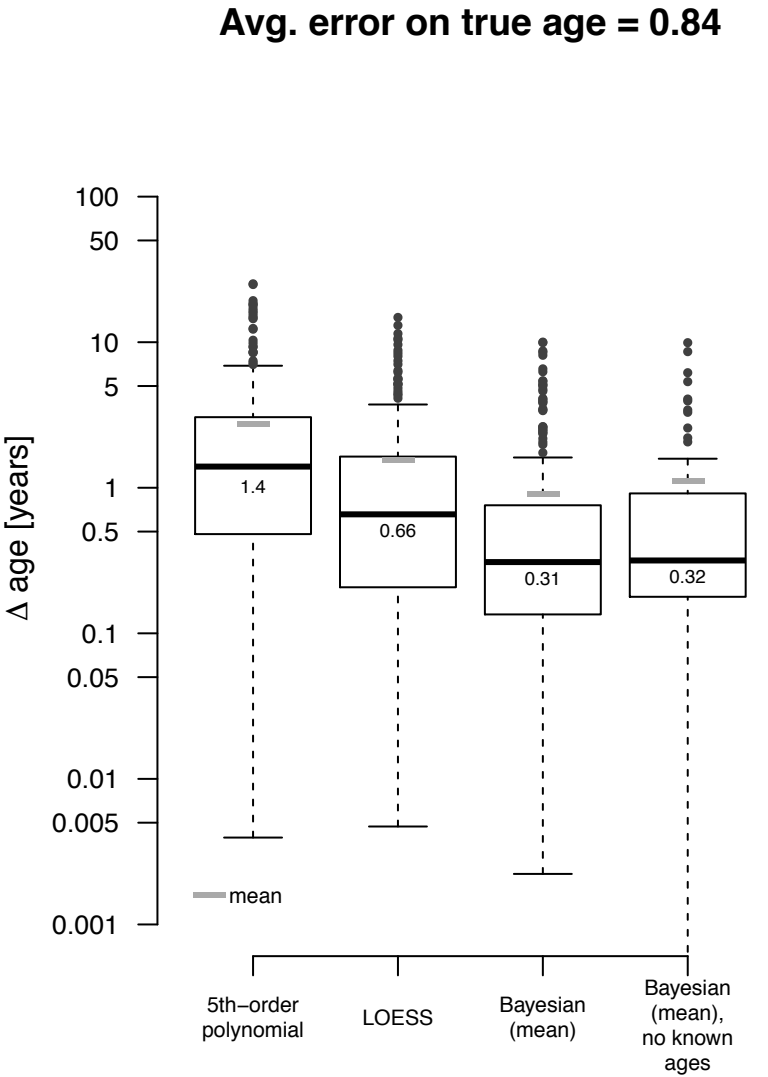

Avg. error on true age = 0.186
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
1.03
0.64
0.26
0.33
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

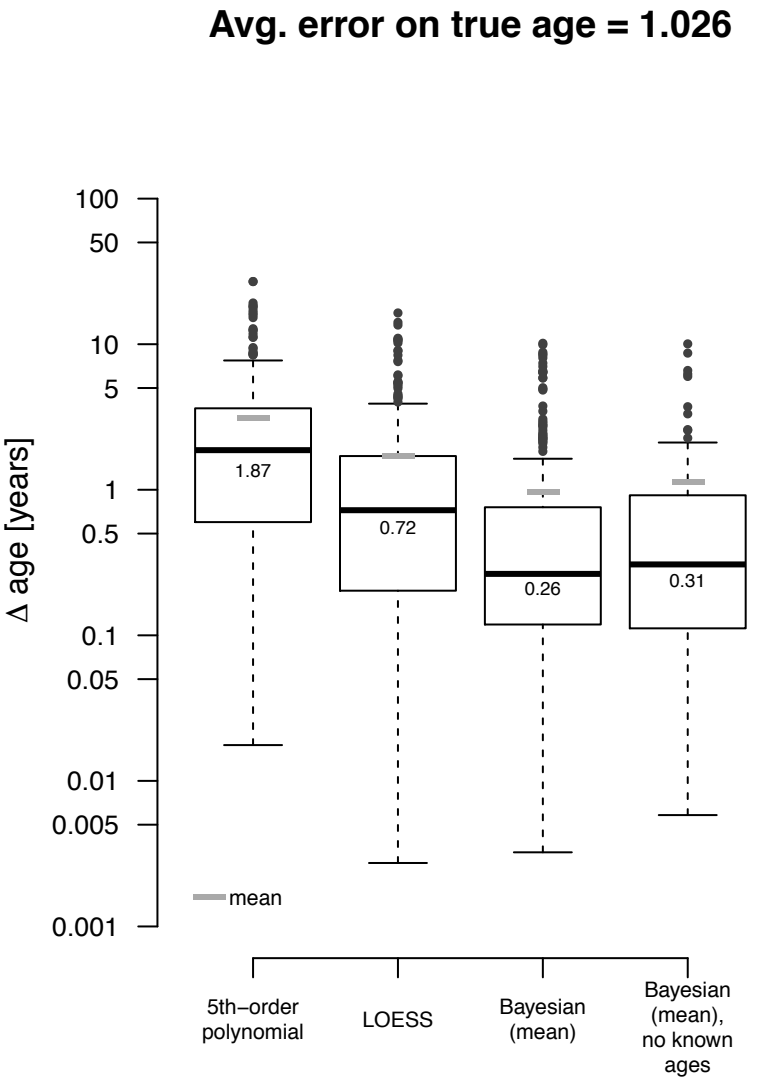

Avg. error on true age = 0.341
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.19
0.62
0.26
0.25
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

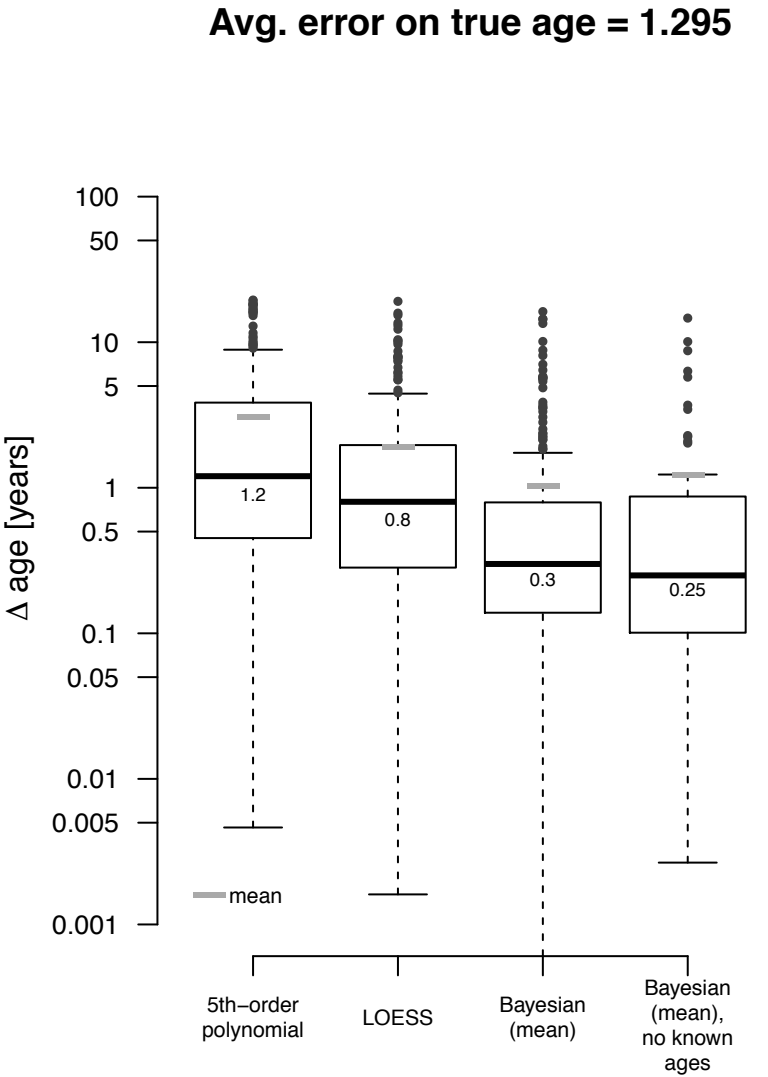
Avg. error on true age = 0.506



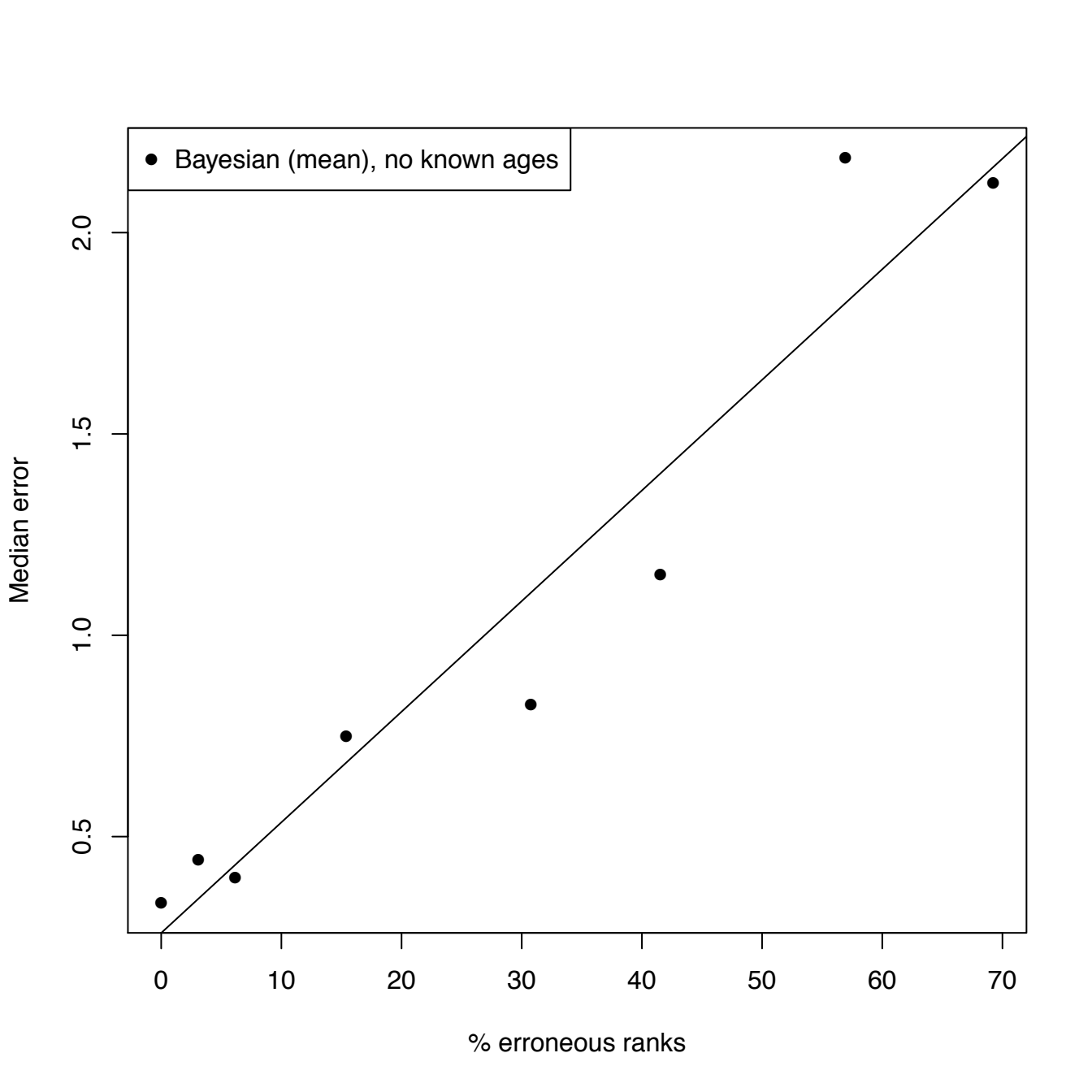

Avg. error on true age = 0.594
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.15
0.57
0.28
0.29
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

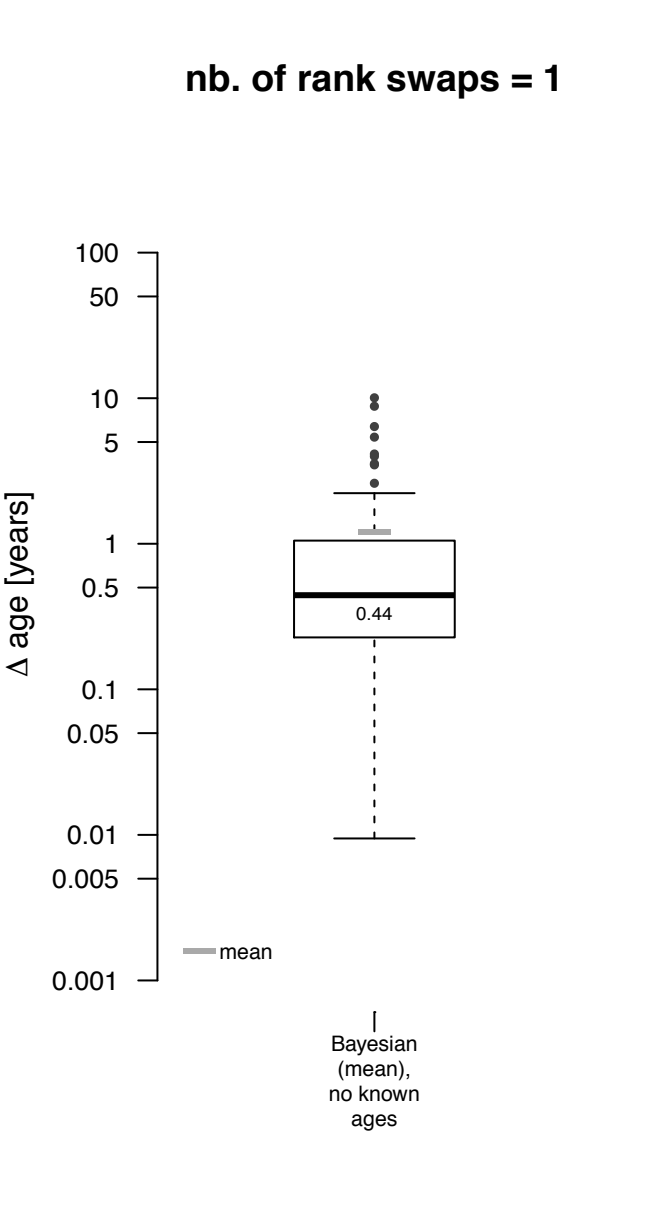

Avg. error on true age = 0.84
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.4
0.66
0.31
0.32
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages


Avg. error on true age = 1.026
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.87
0.72
0.26
0.31
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

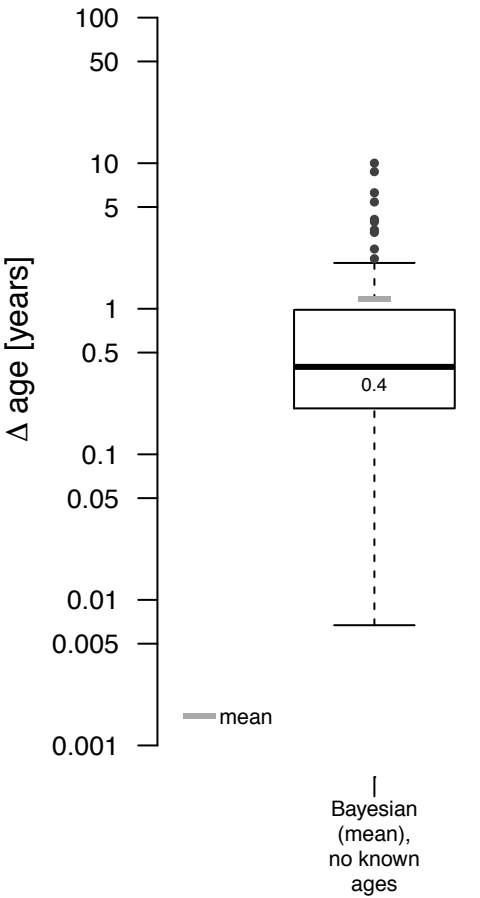

Avg. error on true age = 1.295
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.2
0.8
0.3
0.25
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

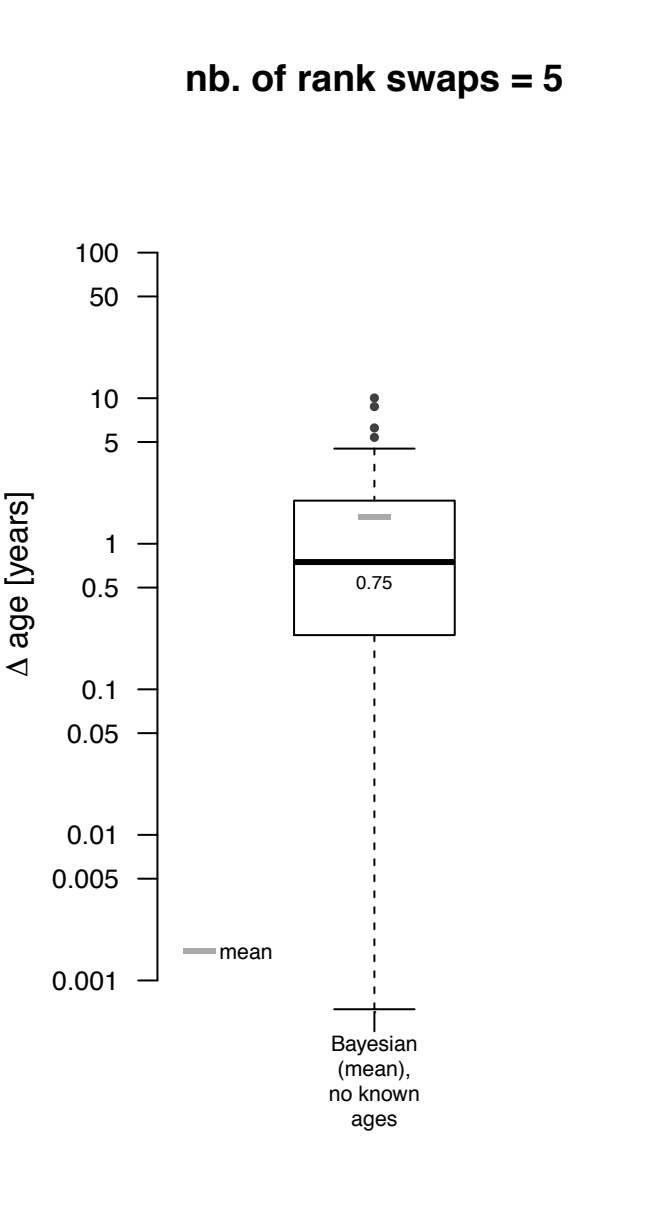
Supplementary Figure S5. *Estimation robustness to error in ranking order.* We repeated validation from Figures 1 and 2, however, introduced different amounts of error in the ranking order (all errors we introduce are consistent with the age brackets). As changing the ranking order would require to adjust the age of the individuals to reflect the altered ranking order, we focus on the performance of our Bayesian method when no ages are considered known. This prevents that the effects of errors in ranking order and age (see Supplementary Figure S4) are conflated. Panel A summarizes the results showing the medians of the differences between estimated and actual ages of the individuals, Panel B gives the corresponding distributions in form of boxplots.


Bayesian (mean), no known ages
Median error
0.5
1.0
1.5
2.0
0
10
20
30
40
50
60
70
% erroneous ranks

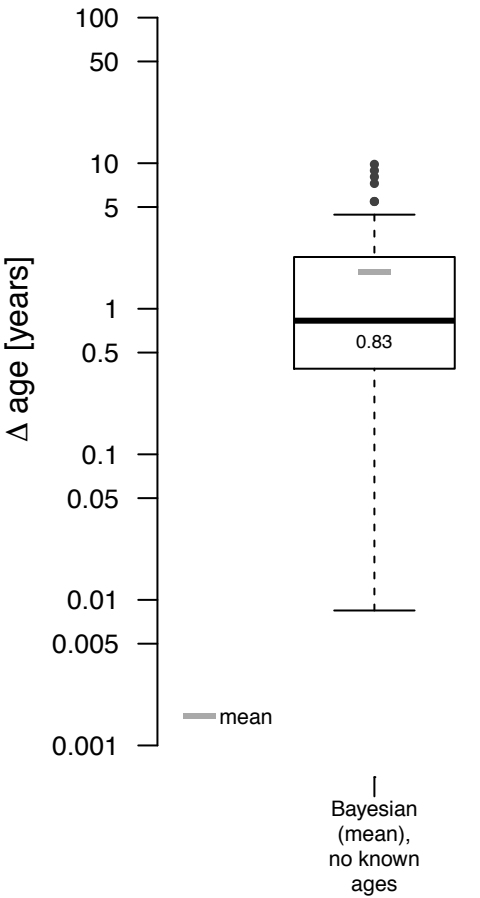

nb. of rank swaps = 1
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
0.44
mean
Bayesian
(mean),
no known
ages

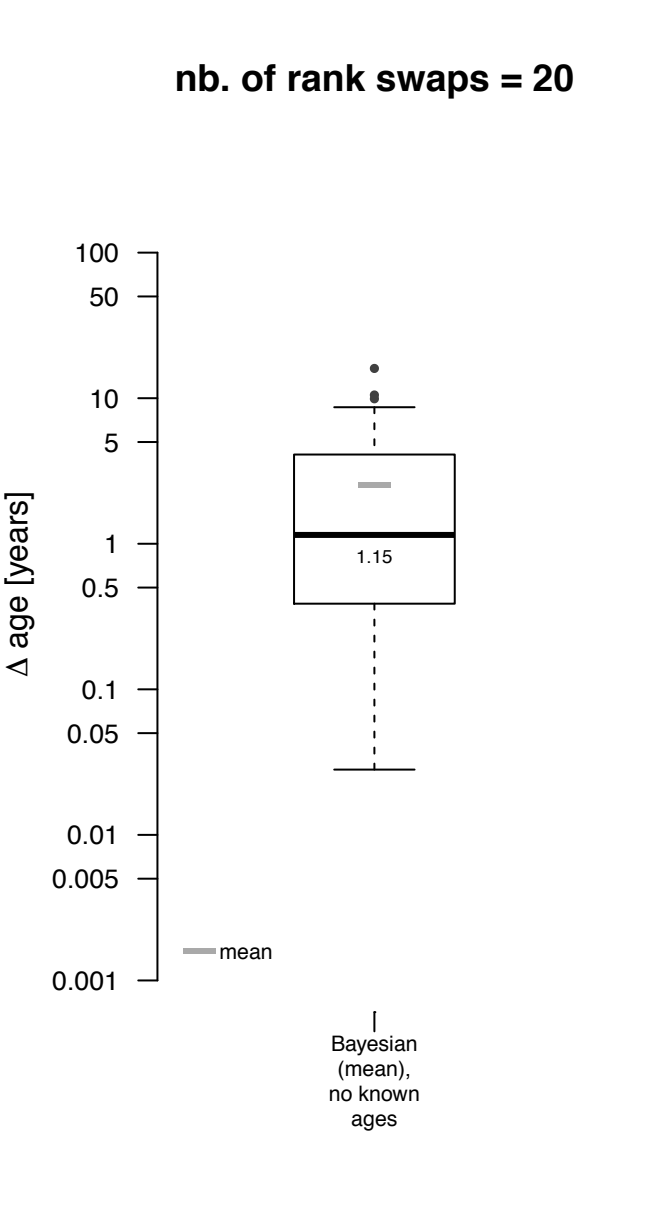
nb. of rank swaps = 2

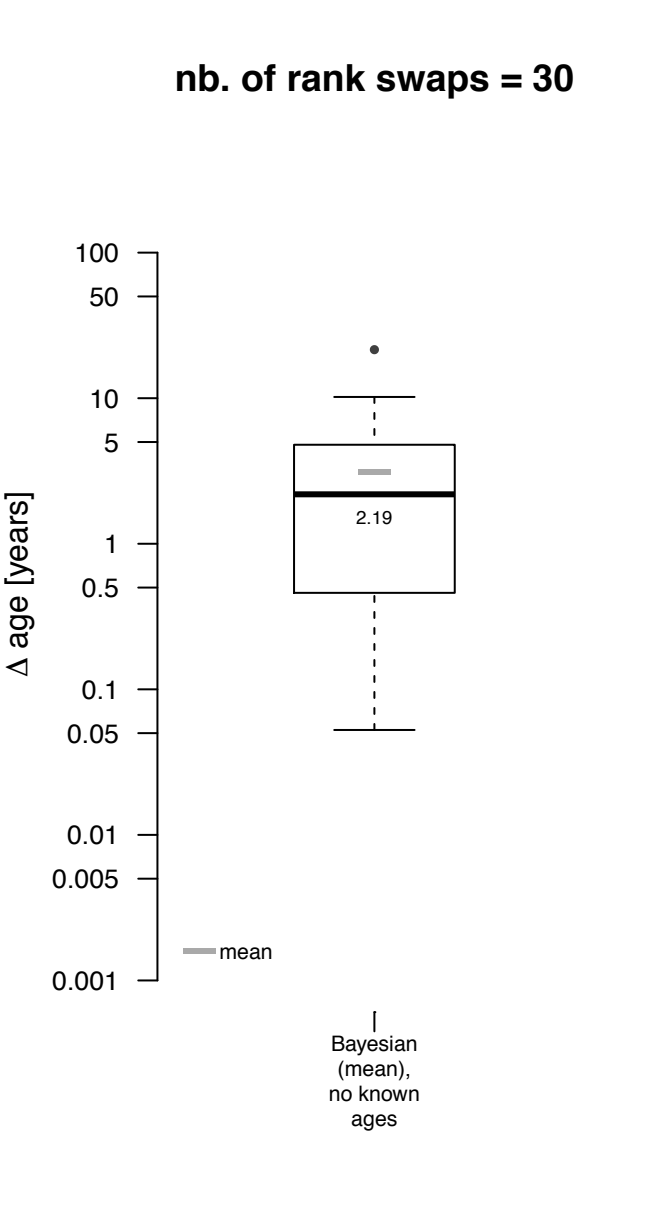



nb. of rank swaps = 5
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
0.75
mean
Bayesian
(mean),
no known
ages

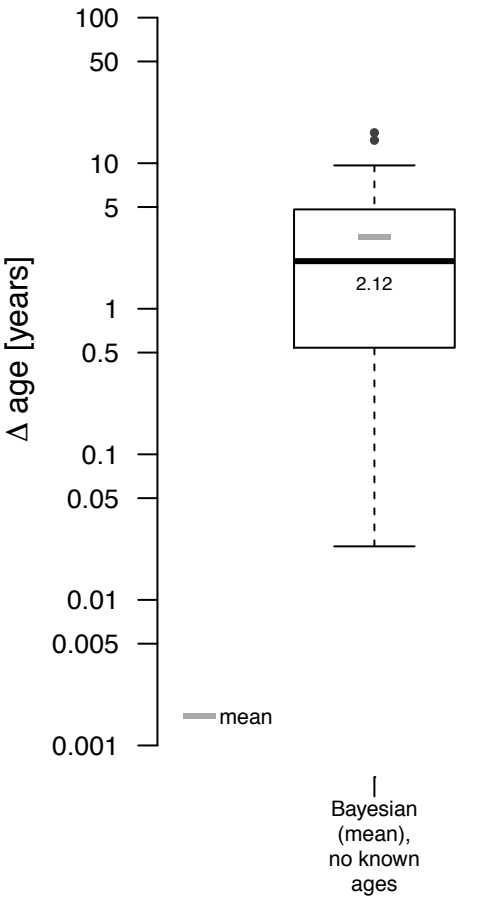
nb. of rank swaps = 10



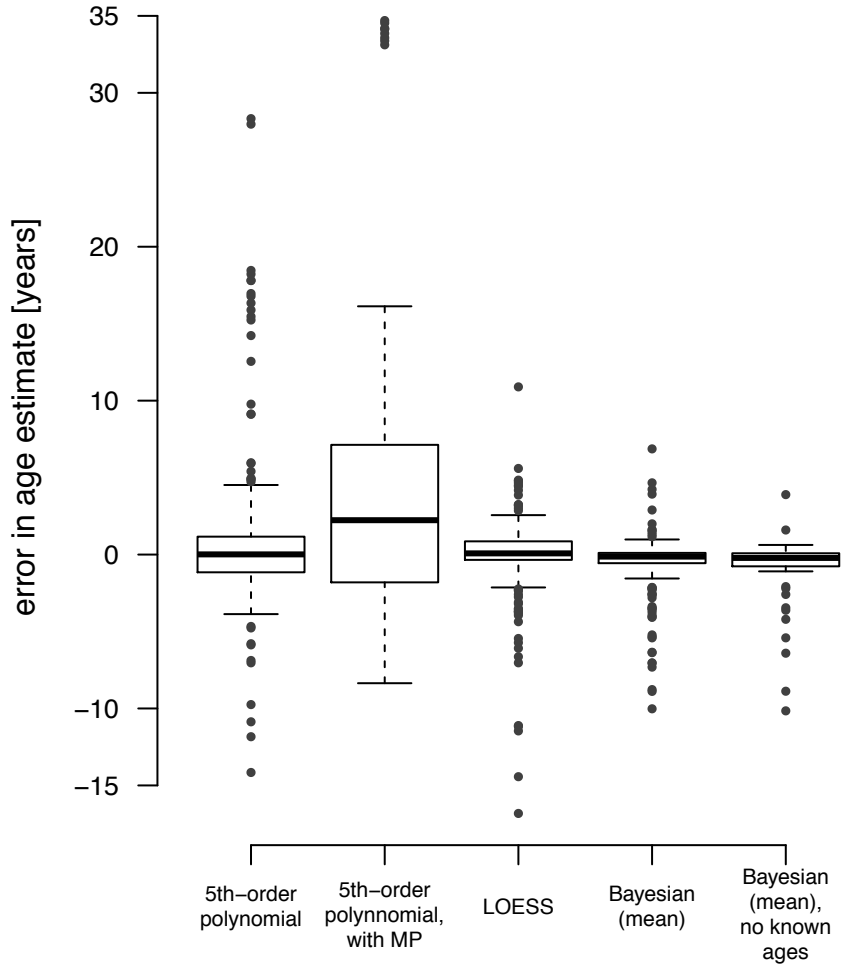

nb. of rank swaps = 20
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
1.15
mean
Bayesian
(mean),
no known
ages


nb. of rank swaps = 30
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
2.19
mean
Bayesian
(mean),
no known
ages

nb. of rank swaps = 50



Supplementary Figure S6. *Raw values behind Figure 1.* We show the same distributions as in Figure 1 in the main text, however, without showing absolute differences and with a y-axis in natural scale.


error in age estimate [years]
35
30
20
10
0
−10
−15
5th−order polynomial
5th−order polynnomial, with MP
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

SUP. TABLES

Supplementary Table S1. *Numerical values corresponding to absolute differences between actual and estimated ages shown in Figure 1 of the main text.* The minimum, 25th percentile, median, mean, 75th percentile and maximum given in the last row (total) directly correspond to the boxplots plotted in Figure 1. The remaining rows provide more detail as the results are split by age cohort. Bold red values indicate worst, bold black best performance. See legend of Figure 1 and explanation of the benchmarking procedure in the main text for further information. Note that photographs in the Headland database (13) were taken in different years (between 1972 and 2010), and all ages and age estimates were therefore adjusted to the present day (2015). Hence, the youngest age is 15 explaining why the 10-20 cohort is the first row.
Abbreviations: minimum (min.), maximum (max.), percentile (per.), standard deviation (sd.), mid-point (MP)

Age Cohort	Sample Size	Statis-tic	5th-order polynomial	5th-order polynomial, with MP	LOESS	Gibbs (mean)	Gibbs (mean), no known ages
10-20	10	min.	0.06	**1.17**	0.03	**0.00**	0.04
		25th per.	0.71	**6.71**	0.12	**0.08**	0.15
		median	1.64	**7.92**	**0.18**	**0.18**	0.24
		mean (sd.)	4.71 (6.30)	**7.96** (3.20)	**0.25** (0.20)	**0.25** (0.27)	0.28 (0.23)
		75th per.	3.91	**8.88**	**0.29**	0.31	0.32
		max.	**18.46**	14.55	**0.84**	1.12	0.85
20-45	40	min.	0.01	**0.02**	0.01	**0.00**	**0.00**
		25th per.	0.23	**1.70**	0.20	**0.11**	0.12
		median	0.67	**4.32**	0.57	**0.25**	0.26
		mean (sd.)	1.13 (1.19)	**4.25** (2.73)	0.94 (1.11)	**0.45** (0.58)	0.47 (0.56)
		75th per.	1.65	**6.23**	1.19	**0.49**	0.53
		max.	5.41	**10.37**	5.59	3.38	**2.59**
45+	15	min.	0.08	0.07	0.14	**0.01**	**0.21**
		25th per.	1.64	**1.81**	1.25	**0.57**	0.88
		median	3.42	2.89	2.71	**1.03**	**3.46**
		mean (sd.)	5.38 (6.01)	**8.02** (11.03)	4.09 (4.05)	**2.57** (2.68)	3.45 (3.15)
		75th per.	5.94	**11.83**	5.28	**4.06**	4.81
		max.	28.32	**34.70**	16.82	**10.02**	10.15
Total	65	min.	0.01	**0.02**	0.01	**0.00**	**0.00**
		25th per.	0.38	**2.05**	0.21	**0.12**	0.18
		median	1.16	**4.39**	0.64	**0.29**	0.33
		mean (sd.)	2.66 (4.35)	**5.69** (6.09)	1.47 (2.36)	**0.91** (1.64)	1.13 (2.01)
		75th per.	2.90	**7.36**	1.57	**0.80**	0.90
		max.	28.32	**34.70**	16.82	**10.02**	10.15

Supplementary Table S2. *Kolmogorov-Smirnov p-values and Bayes factors for all pairwise comparisons of error distributions shown in Figure 1.* BFs greater than three are considered positive evidence, above 150 as strong evidence. *Abbreviations:* mid-point (MP), Bayes factor (BF)

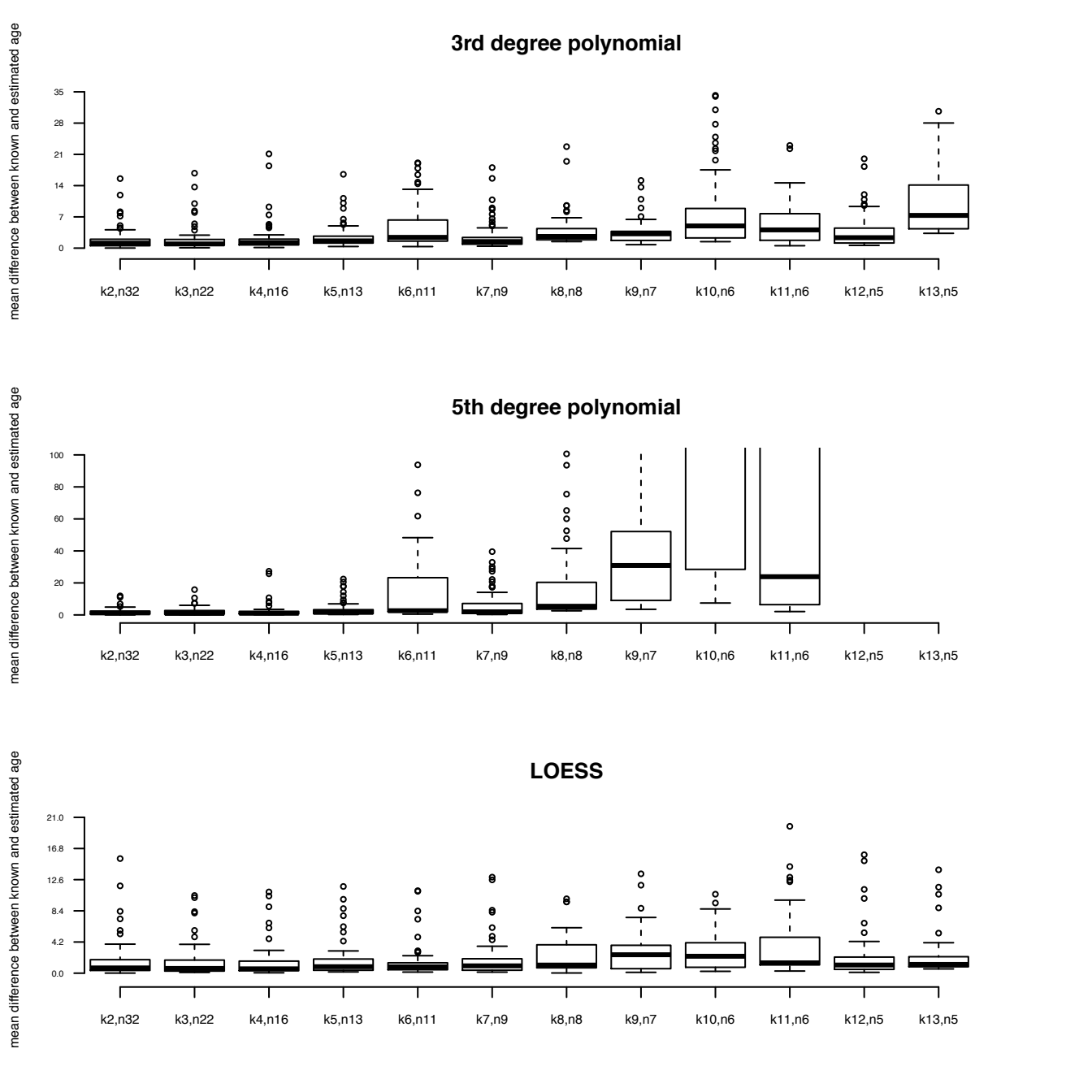
	5th-order polynomial	**5th-order polynomial, with MP**	**LOESS**	**Gibbs (mean)**
5th-order polynomial, with MP	*p*=1.554312e-15; BF=4.128145e+20			
LOESS	*p*=0.0004320986; BF= 29.39566	*p*=1.776357e-15; BF= 7.426503e+38		
Gibbs (mean)	*p*=3.108624e-15; BF=2.81064e+12	*p*=1.554312e-15; BF=2.04265e+63	*p*=4.486276e-06; BF=1377.745	
Gibbs (mean), no known ages	*p*=9.447281e-06; BF=419.9913	*p*=7.771561e-16; BF=6.354463e+27	*p*=0.02777288; BF=1.054143	*p*=0.6081314; BF=0.2328565

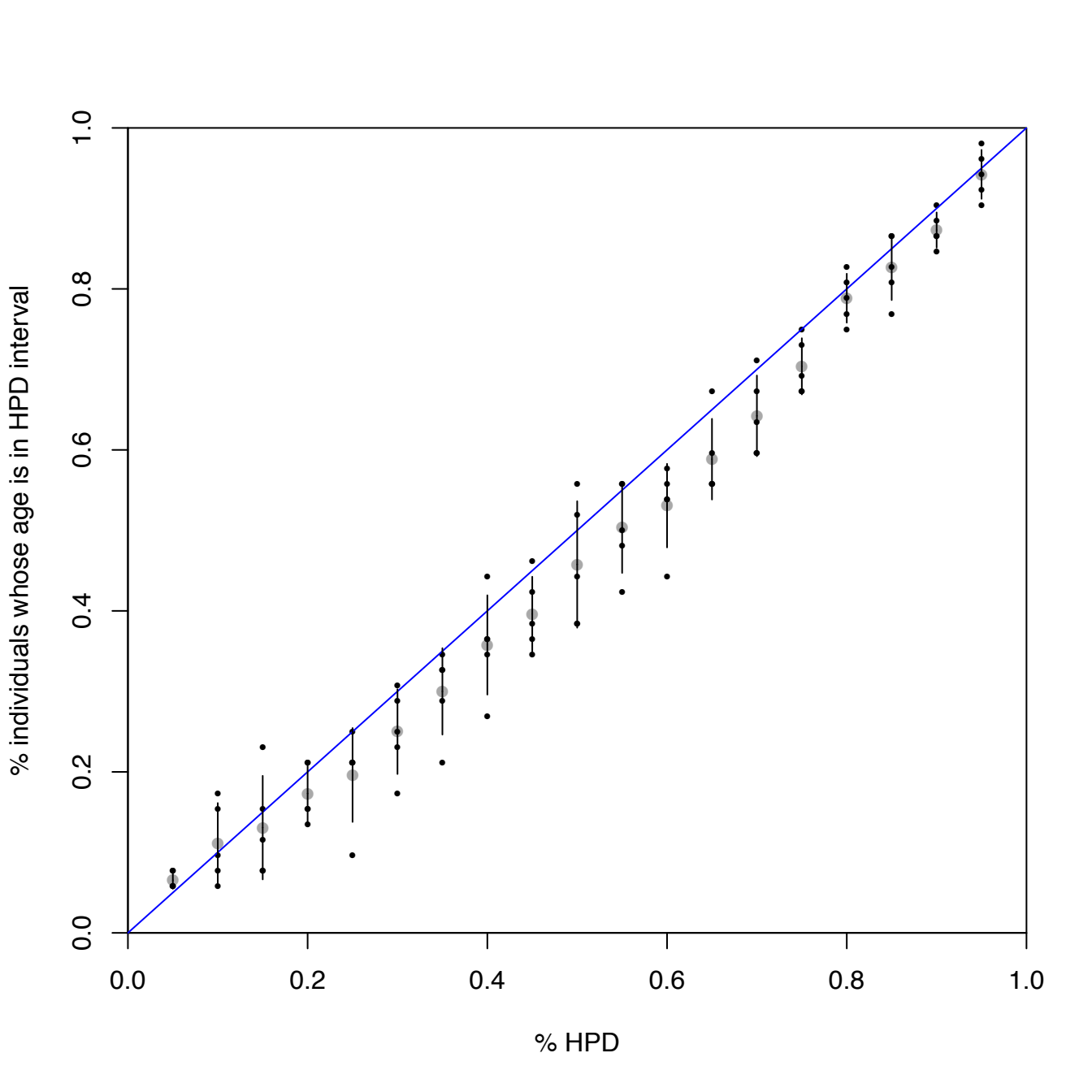
Supplementary Table S3. *Average difference between upper and lower bound of the age bracket and number of accurately known ages for different age cohorts of the Palanan Agta*. For the purposes of this table, the mean value of the upper and lower bound was considered an individual's age and used for grouping into cohorts. Number of exact birth dates and birth dates accurate within +/- 1 year are also displayed.

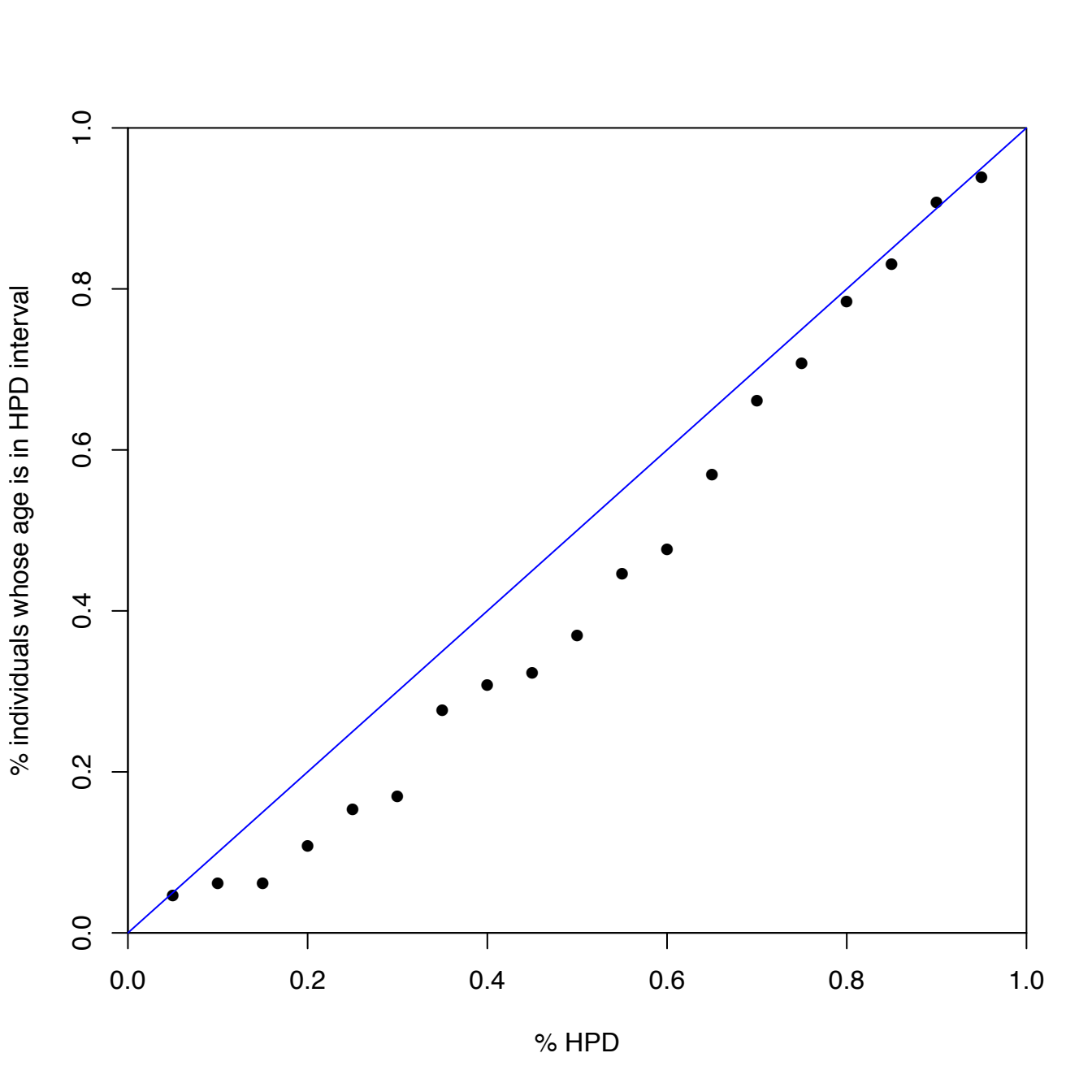
Age Cohort	Sample Size	Average Difference	Number of Exact Birthdates	Percentage of Exact Birthdates	Number of Birthdates +/- 1 year	Percentage of Birthdates +/- 1 year
<1	20	0.16	15	75%	20	100%
1-5	103	1.73	30	29.13%	67	65.05%
5-10	103	3	19	18.45%	33	32.04%
10-20	116	4.1	13	11.21%	33	28.45%
20-45	164	9.47	18	10.98%	26	15.85%
45+	81	18.56	3	3.7%	12	14.81%
Total	587	6.85	98	16.7%	191	32.54%

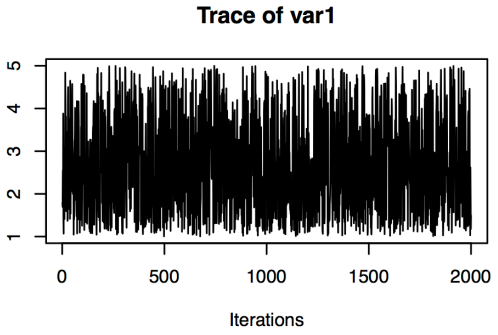
REFERENCES

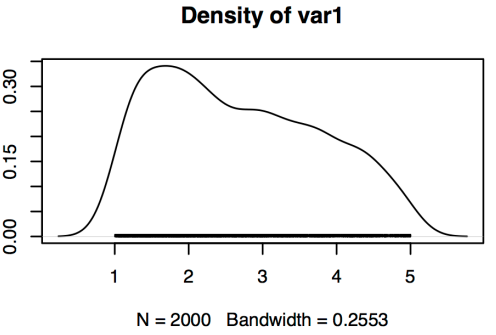
1. Kaplan H, Hill J, Lancaster J, Hurtado A M, Hill K I M, Lancaster J, Hurtado A M (2000) A theory of human life history evolution: diet, intelligence, and longevity. *Evolutionary Anthropology 9*:156–185.
2. Hill K, Hurtado A M (1996) *Aché Life History: The Ecology and Demography of a Foraging People* (Aldine de Gruyter, New Haven).
3. Walsh B (2004) Markov Chain Monte Carlo and Gibbs Sampling. *Lecture Notes for EEB*:1–24.
4. Headland T N, Headland J D, Uehara R T (2011) *Agta Demographic Database: Chronicle of a hunter-gatherer community in transition* (SIL Language and Culture Documentation and Description, 2).
5. Python Software Foundation. (2016) Python Language Reference.
6. R Core Team (2012) R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/
7. Cleveland W S, Grosse E, Shyu W M (1992) Local Regression Models. *Statistical Models in S,* eds Chambers J M, Hastie J J (Wadsworth & Brooks, Pacific Grove, CA), pp 309–376.
8. Gelman A, Rubin DB (1992) Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science 7*(4):457–511.
9. Raftery AE, Lewis SM (1995) The number of iterations, convergence diagnostics and generic Metropolis algorithms. In: Gilks WR, Spiegelhalter DJ, Richardson S, eds. *Practical Markov Chain Monte Carlo*. London, UK: Chapman and Hall; pp 1–15.

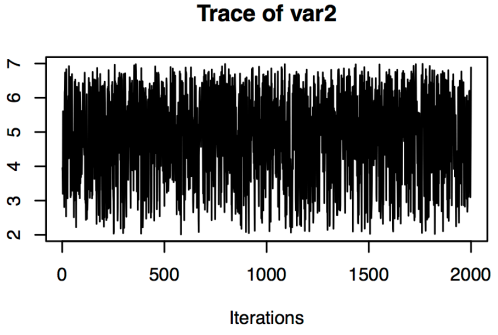

3rd degree polynomial
mean difference between known and estimated age
0
7
14
21
28
35
k2,n32
k3,n22
k4,n16
k5,n13
k6,n11
k7,n9
k8,n8
k9,n7
k10,n6
k11,n6
k12,n5
k13,n5
5th degree polynomial
mean difference between known and estimated age
0
20
40
60
80
100
k2,n32
k3,n22
k4,n16
k5,n13
k6,n11
k7,n9
k8,n8
k9,n7
k10,n6
k11,n6
k12,n5
k13,n5
LOESS
mean difference between known and estimated age
0.0
4.2
8.4
12.6
16.8
21.0
k2,n32
k3,n22
k4,n16
k5,n13
k6,n11
k7,n9
k8,n8
k9,n7
k10,n6
k11,n6
k12,n5
k13,n5

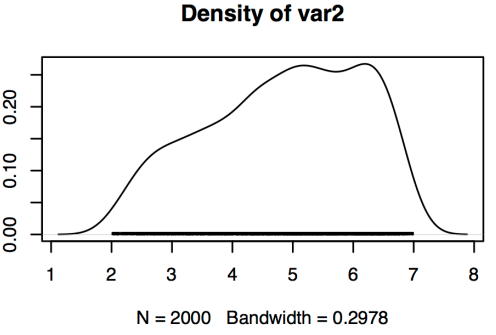

% individuals whose age is in HPD interval
0.0
0.2
0.4
0.6
0.8
1.0
% HPD
0.0
0.2
0.4
0.6
0.8
1.0

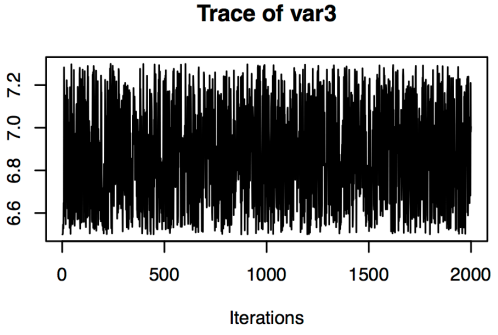

% individuals whose age is in HPD interval
0.0
0.2
0.4
0.6
0.8
1.0
% HPD
0.0
0.2
0.4
0.6
0.8
1.0

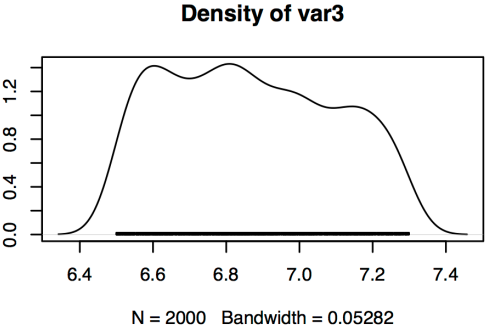

Trace of var1
1
2
3
4
5
0
500
1000
1500
2000
Iterations

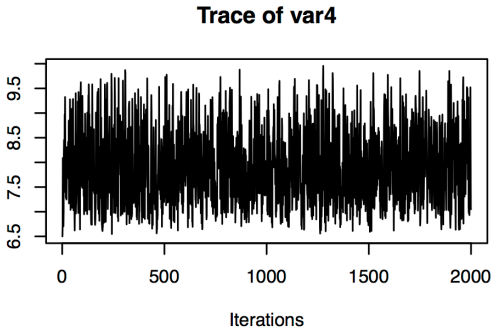

Density of var1
0.00
0.15
0.30
1
2
3
4
5
N = 2000 Bandwidth = 0.2553

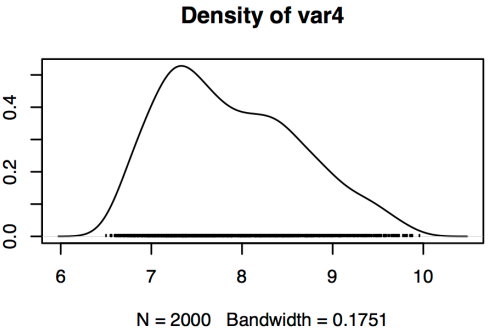

Trace of var2
2
3
4
5
6
7
0
500
1000
1500
2000
Iterations

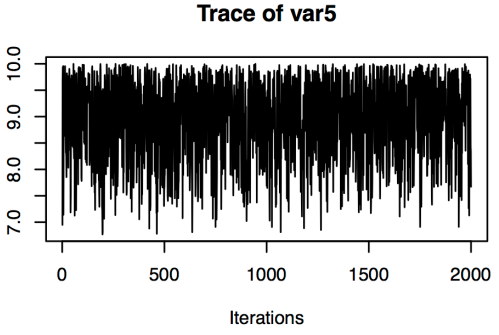

Density of var2
0.00
0.10
0.20
1
2
3
4
5
6
7
8
N = 2000 Bandwidth = 0.2978

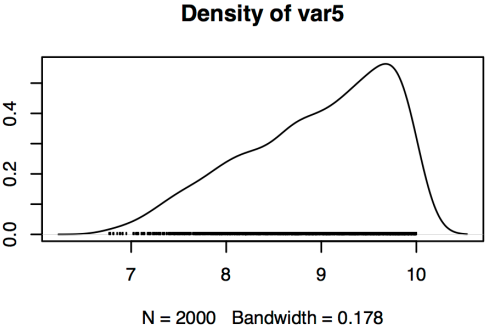

Trace of var3
6.6
6.8
7.0
7.2
0
500
1000
1500
2000
Iterations

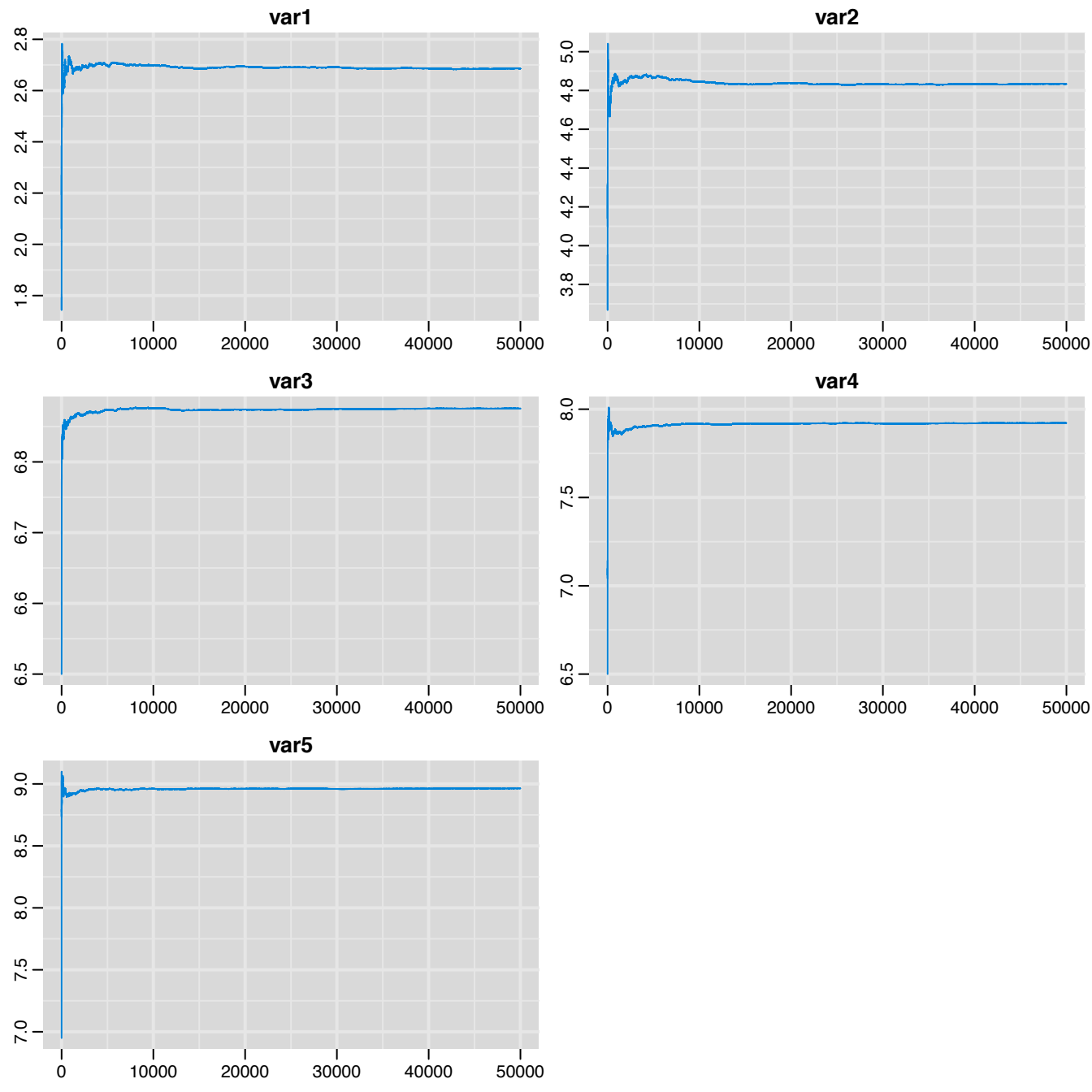

Density of var3
0.0
0.4
0.8
1.2
6.4
6.6
6.8
7.0
7.2
7.4
N = 2000 Bandwidth = 0.05282

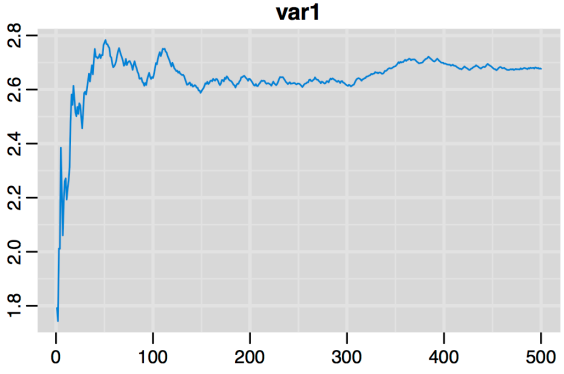

Trace of var4
6.5
7.5
8.5
9.5
0
500
1000
1500
2000
Iterations

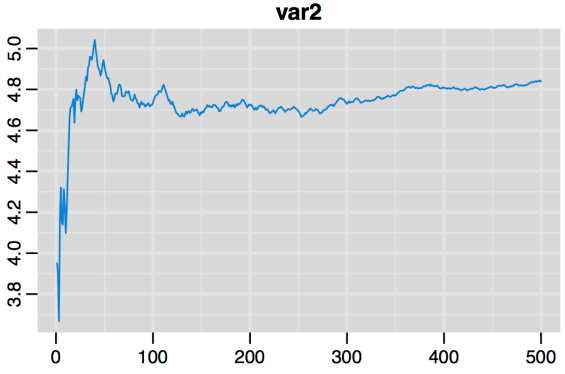

Density of var4
0.0
0.2
0.4
6
7
8
9
10
N = 2000 Bandwidth = 0.1751

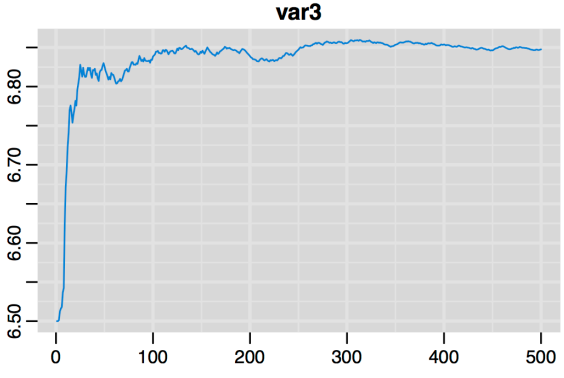

Trace of var5
7.0
8.0
9.0
10.0
0
500
1000
1500
2000
Iterations

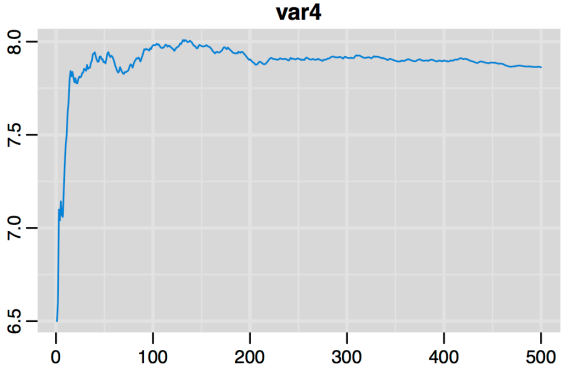

Density of var5
0.0
0.2
0.4
7
8
9
10
N = 2000 Bandwidth = 0.178

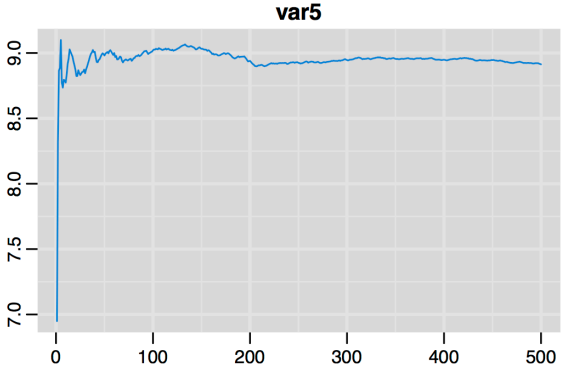

var1
1.8
2.0
2.2
2.4
2.6
2.8
0
10000
20000
30000
40000
50000
var2
3.8
4.0
4.2
4.4
4.6
4.8
5.0
0
10000
20000
30000
40000
50000
var3
6.5
6.6
6.7
6.8
0
10000
20000
30000
40000
50000
var4
6.5
7.0
7.5
8.0
0
10000
20000
30000
40000
50000
var5
7.0
7.5
8.0
8.5
9.0
0
10000
20000
30000
40000
50000

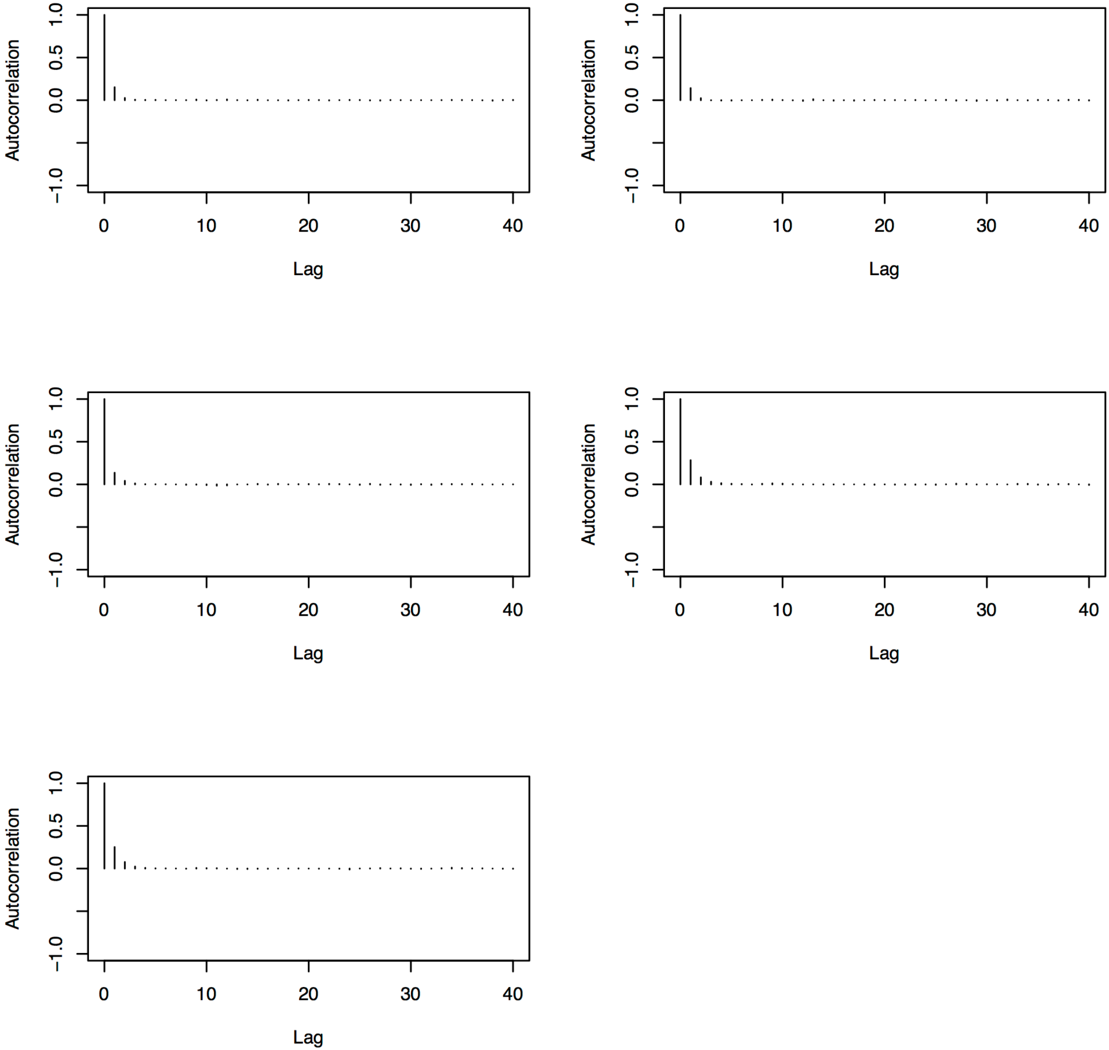

var1
1.8
2.0
2.2
2.4
2.6
2.8
0
100
200
300
400
500


var2
3.8
4.0
4.2
4.4
4.6
4.8
5.0
0
100
200
300
400
500

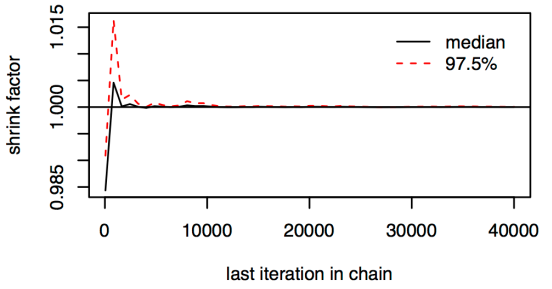

var3
6.50
6.60
6.70
6.80
0
100
200
300
400
500

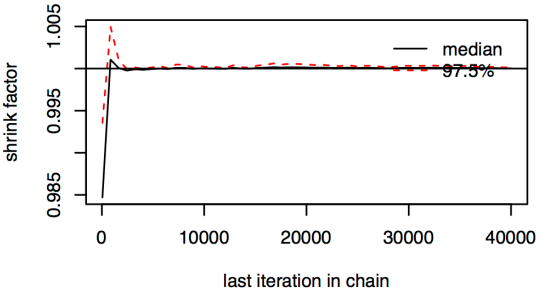

var4
6.5
7.0
7.5
8.0
0
100
200
300
400
500

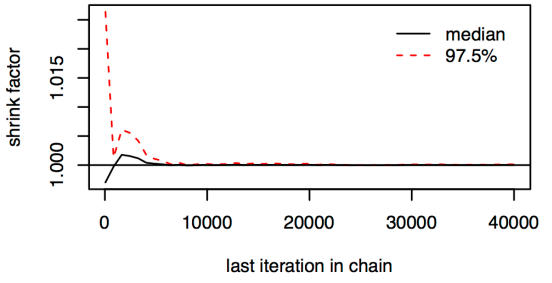

var5
7.0
7.5
8.0
8.5
9.0
0
100
200
300
400
500

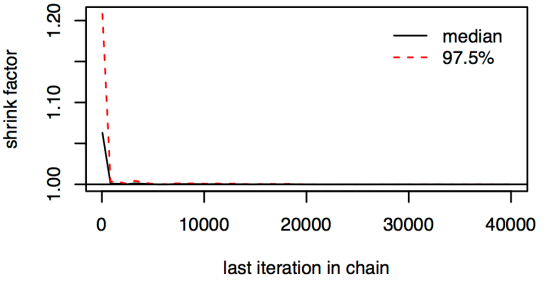


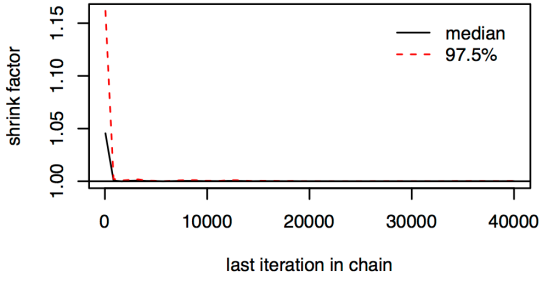
Autocorrelation
1.0
0.5
0.0
−1.0
0
10
20
30
40
Lag
Autocorrelation
1.0
0.5
0.0
−1.0
0
10
20
30
40
Lag
Autocorrelation
1.0
0.5
0.0
−1.0
0
10
20
30
40
Lag
Autocorrelation
1.0
0.5
0.0
−1.0
0
10
20
30
40
Lag
Autocorrelation
1.0
0.5
0.0
−1.0
0
10
20
30
40
Lag

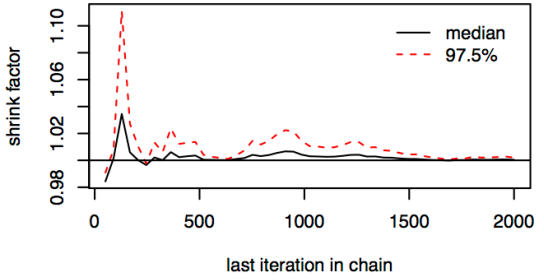

shrink factor
0.985
1.000
1.015
median
97.5%
0
10000
20000
30000
40000
last iteration in chain

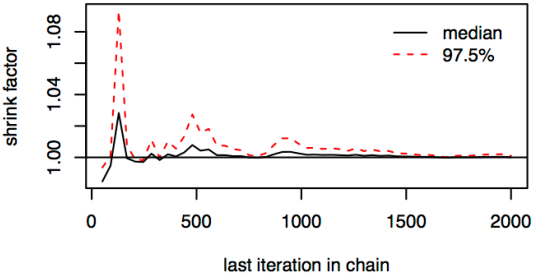

shrink factor
0.985
0.995
1.005
median
97.5%
0
10000
20000
30000
40000
last iteration in chain

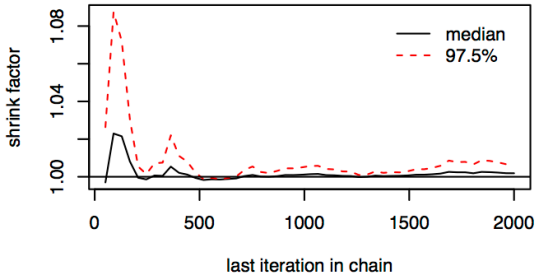

shrink factor
1.000
1.015
median
97.5%
0
10000
20000
30000
40000
last iteration in chain

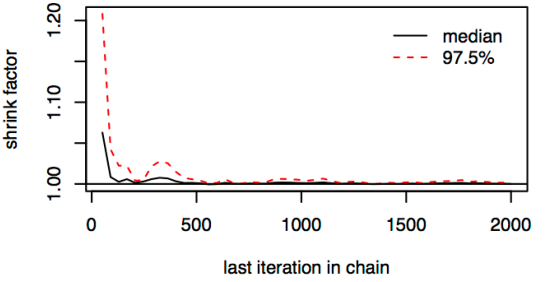

shrink factor
1.00
1.10
1.20
median
97.5%
0
10000
20000
30000
40000
last iteration in chain

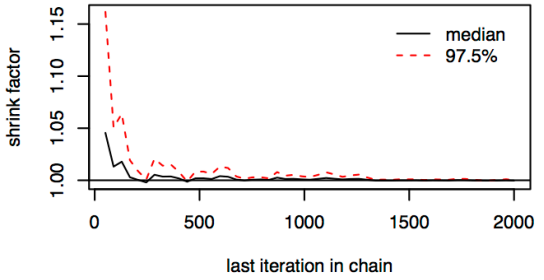

shrink factor
1.00
1.05
1.10
1.15
median
97.5%
0
10000
20000
30000
40000
last iteration in chain

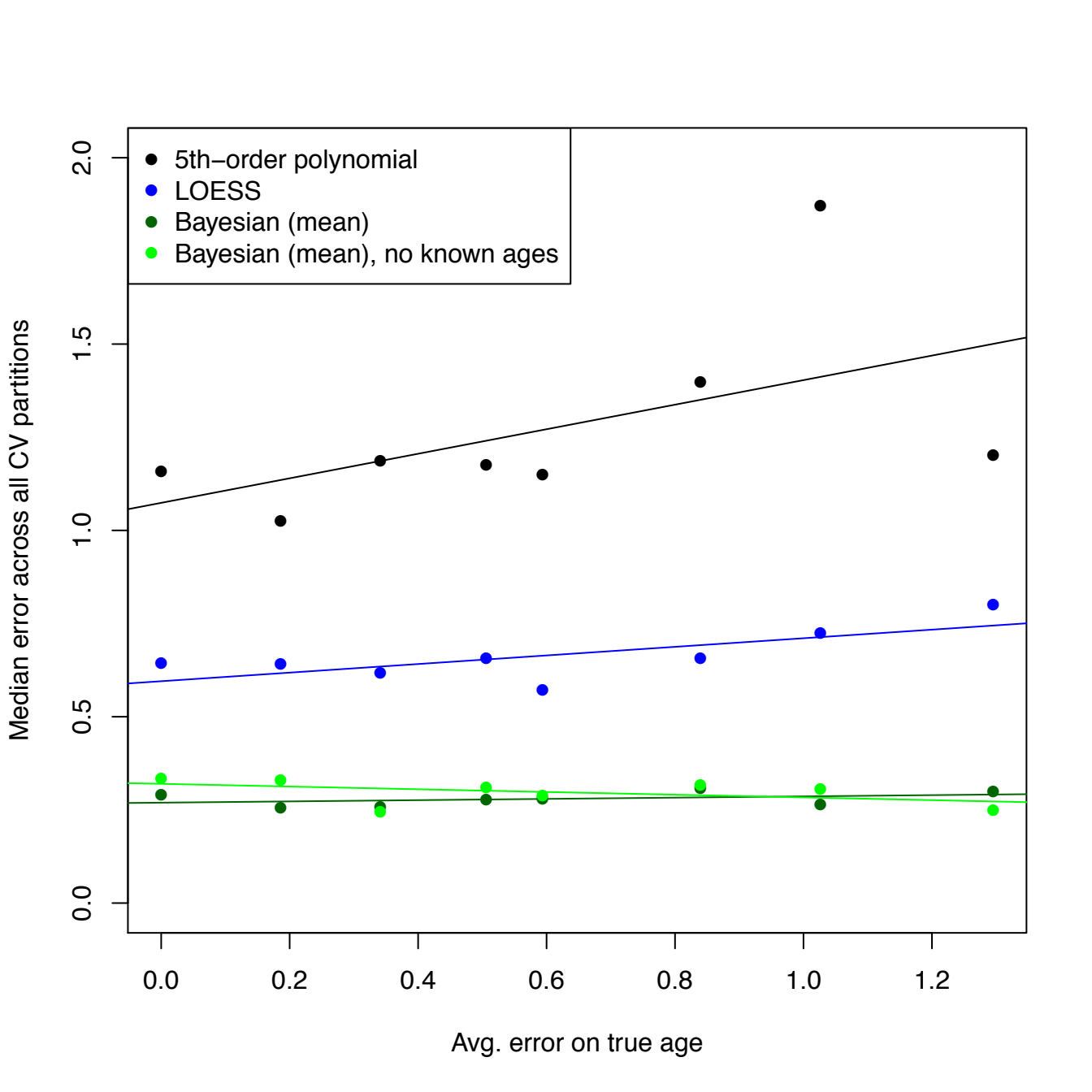

shrink factor
0.98
1.02
1.06
1.10
median
97.5%
0
500
1000
1500
2000
last iteration in chain

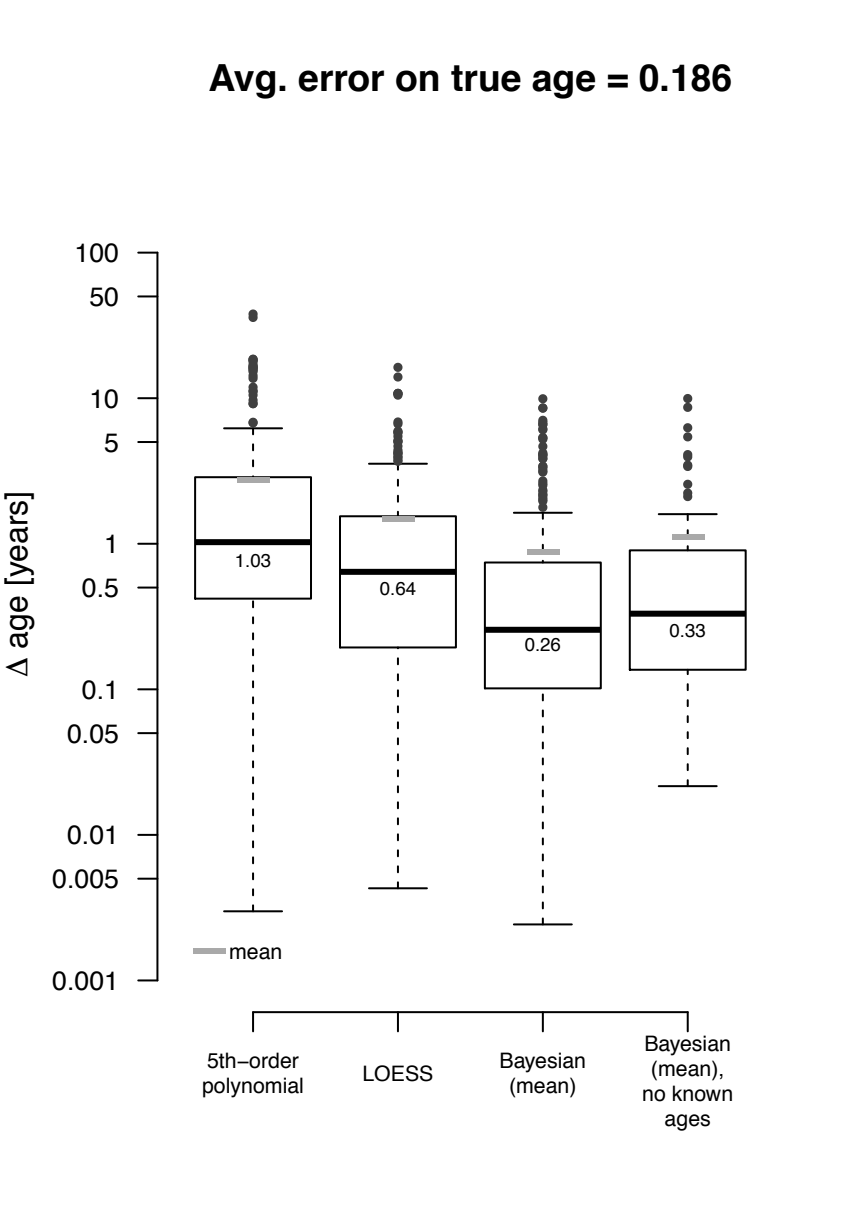

shrink factor
1.00
1.04
1.08
median
97.5%
0
500
1000
1500
2000
last iteration in chain

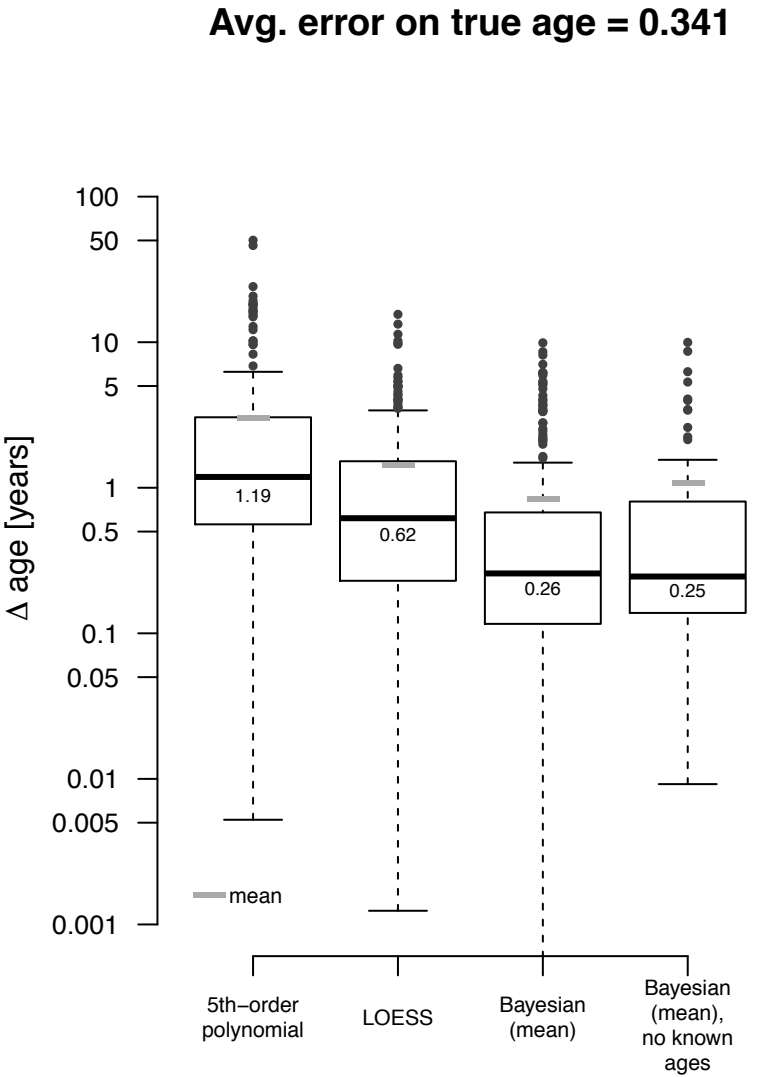

shrink factor
1.00
1.04
1.08
median
97.5%
0
500
1000
1500
2000
last iteration in chain


shrink factor
1.00
1.10
1.20
median
97.5%
0
500
1000
1500
2000
last iteration in chain

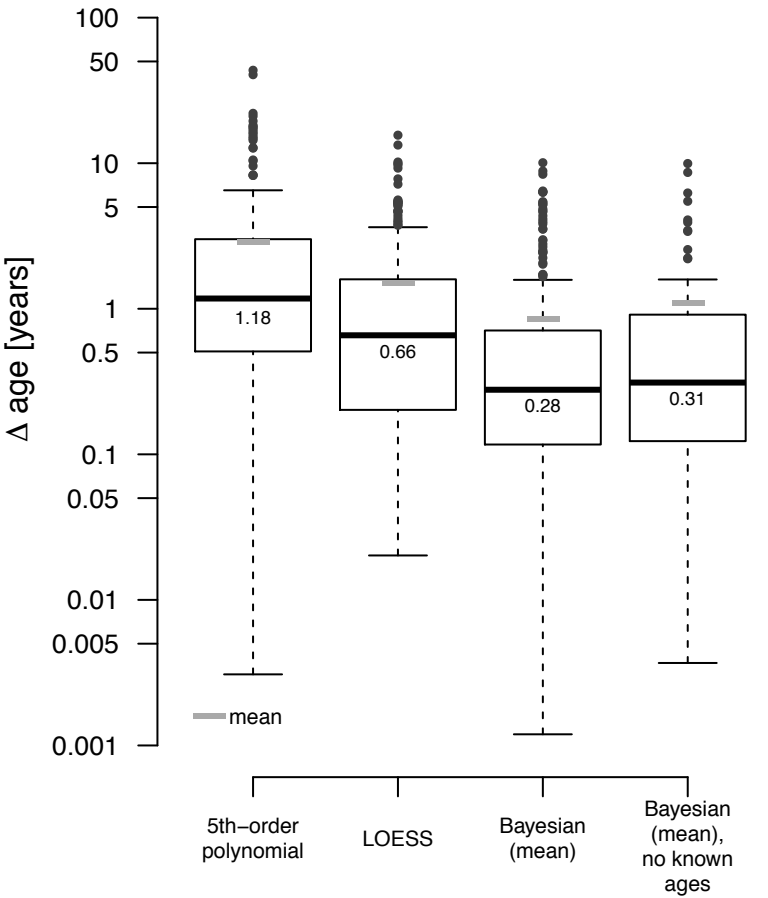

shrink factor
1.00
1.05
1.10
1.15
median
97.5%
0
500
1000
1500
2000
last iteration in chain

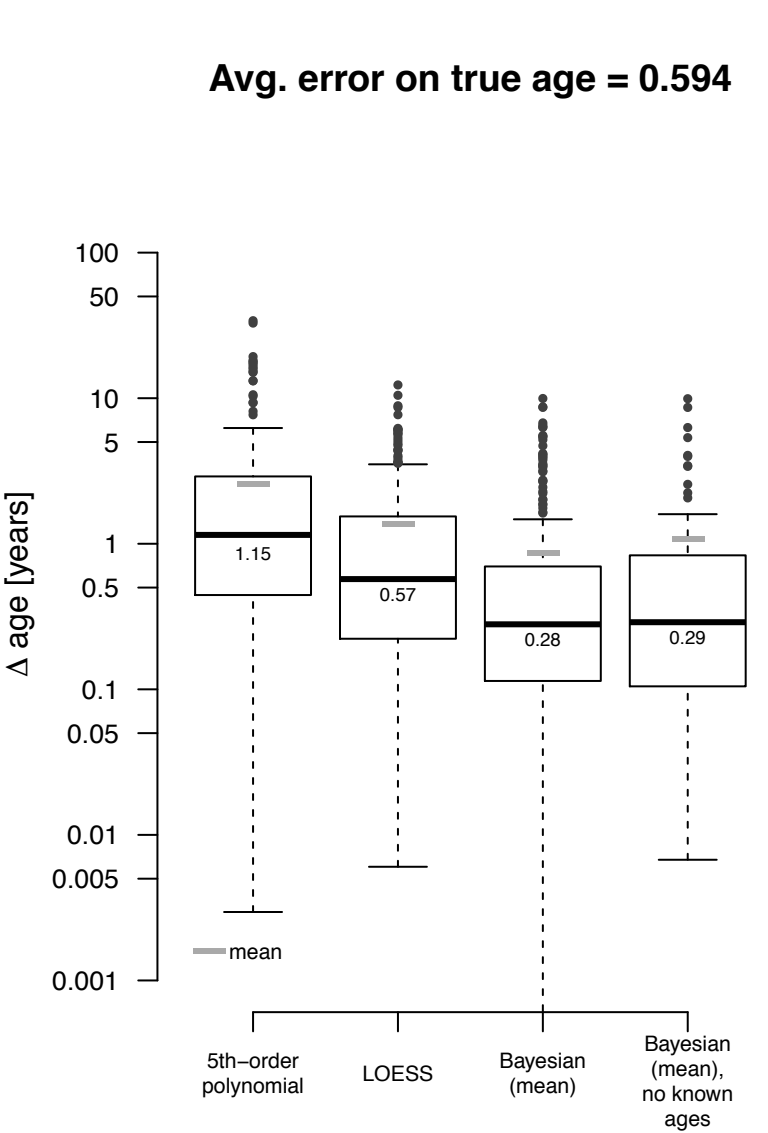

5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages
Median error across all CV partitions
Avg. error on true age
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2

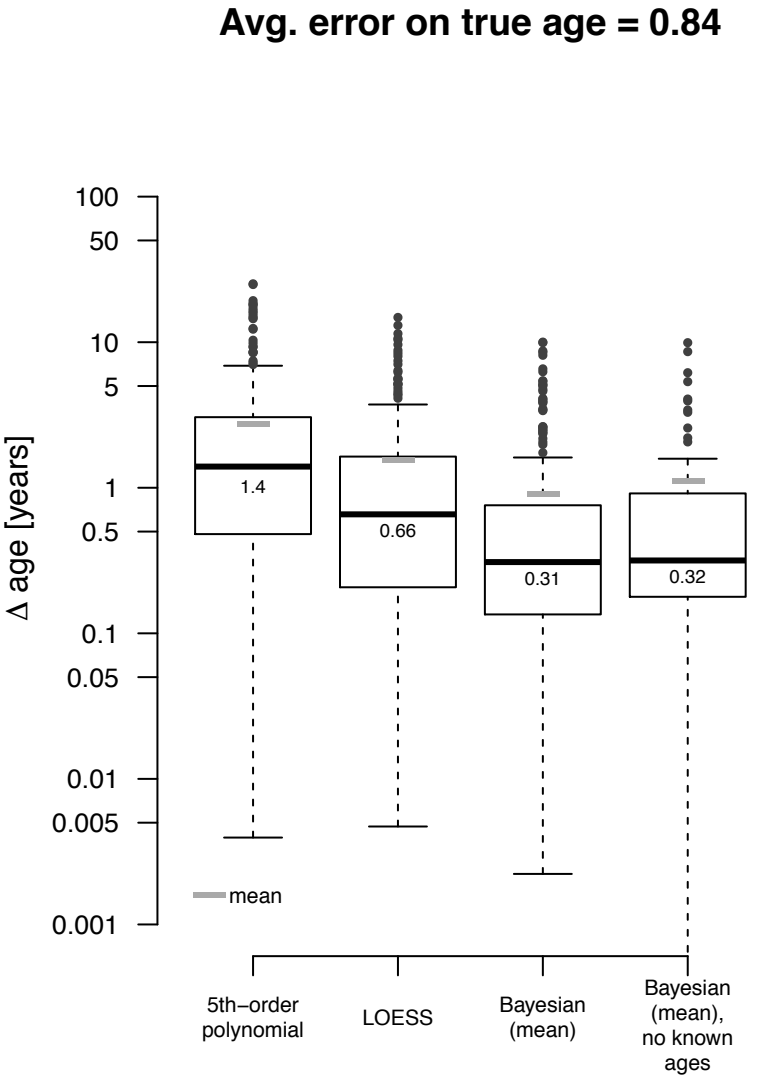

Avg. error on true age = 0.186
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
1.03
0.64
0.26
0.33
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

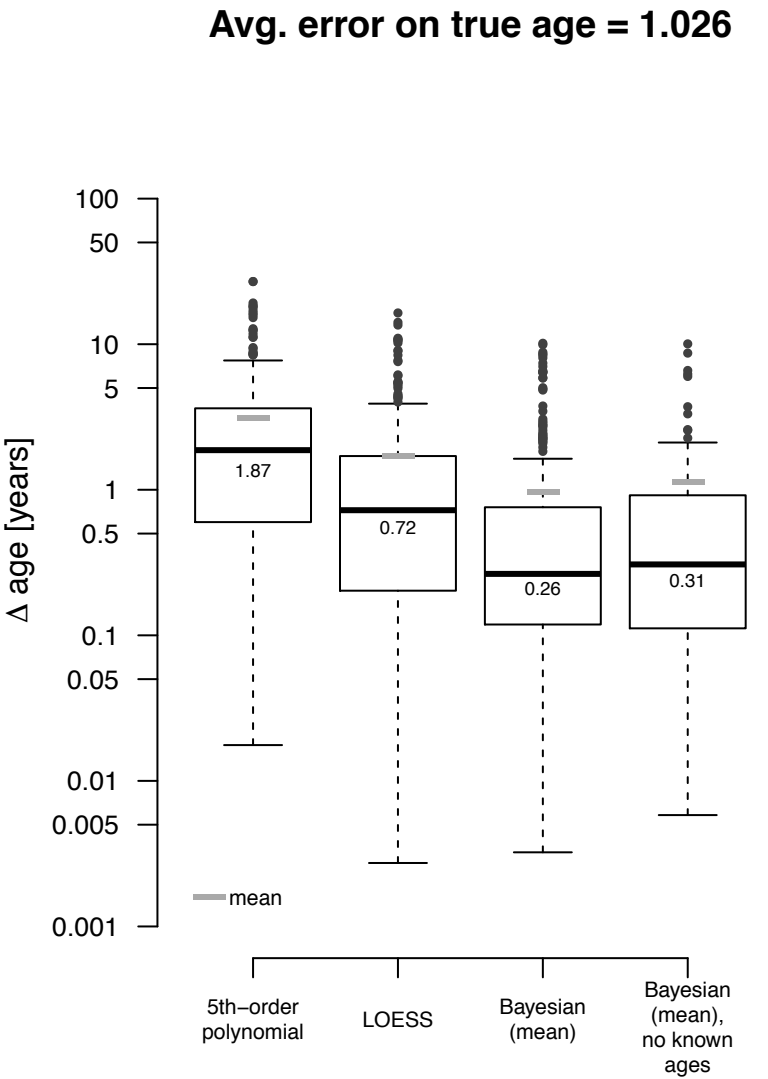

Avg. error on true age = 0.341
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.19
0.62
0.26
0.25
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

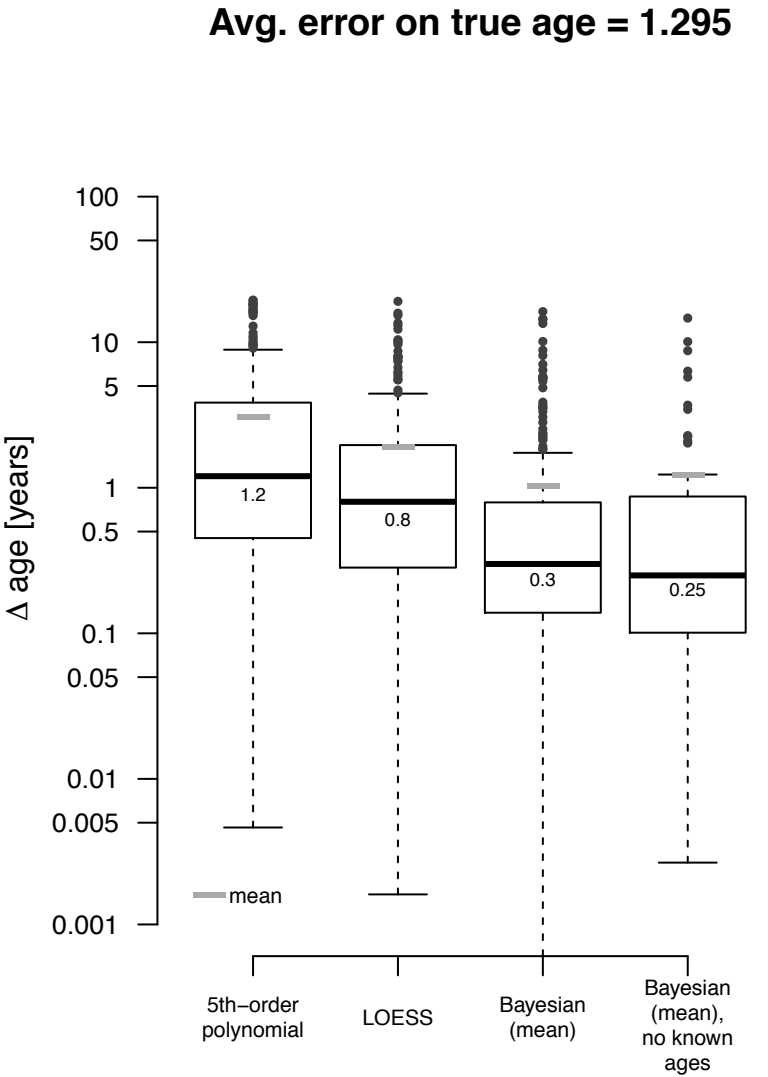
Avg. error on true age = 0.506

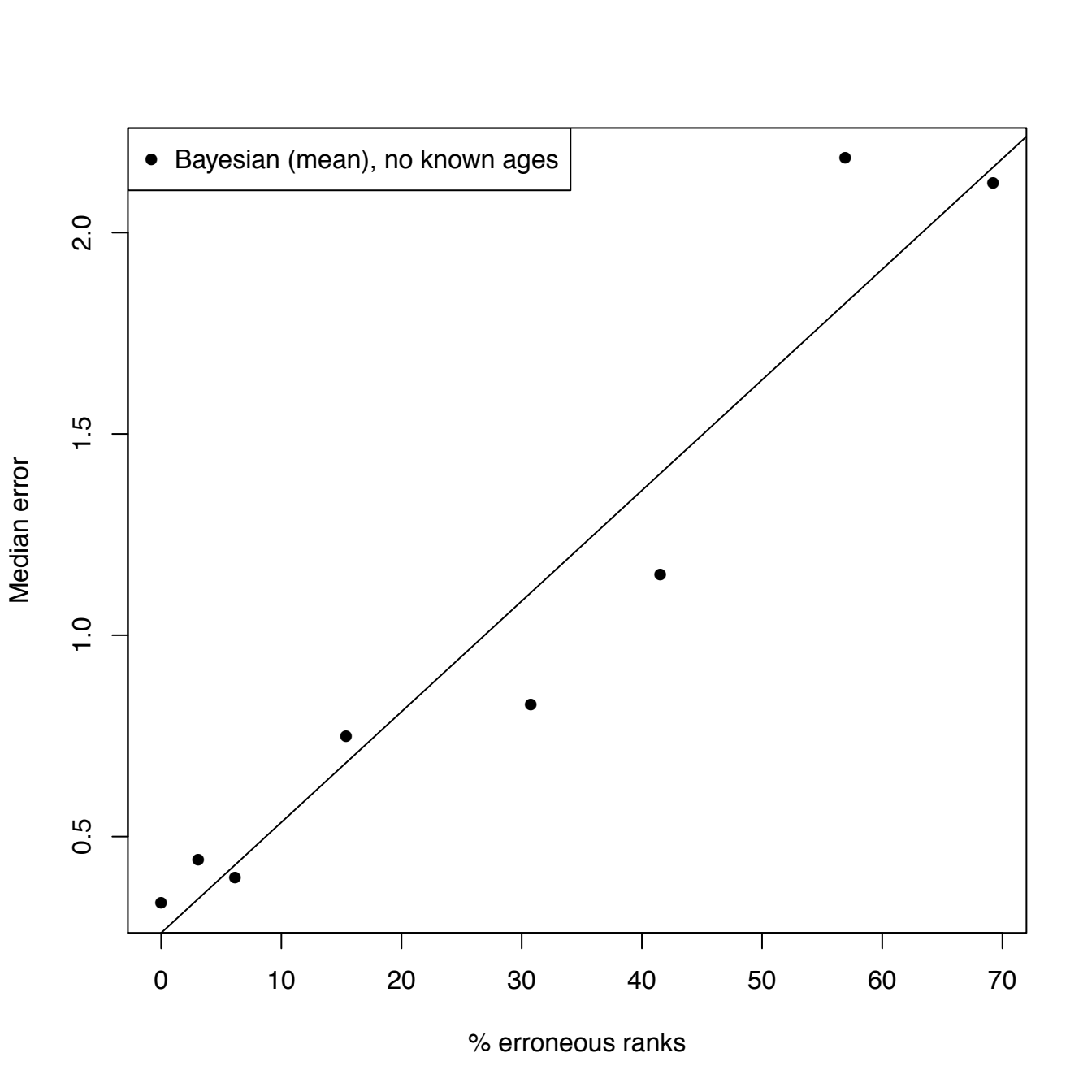


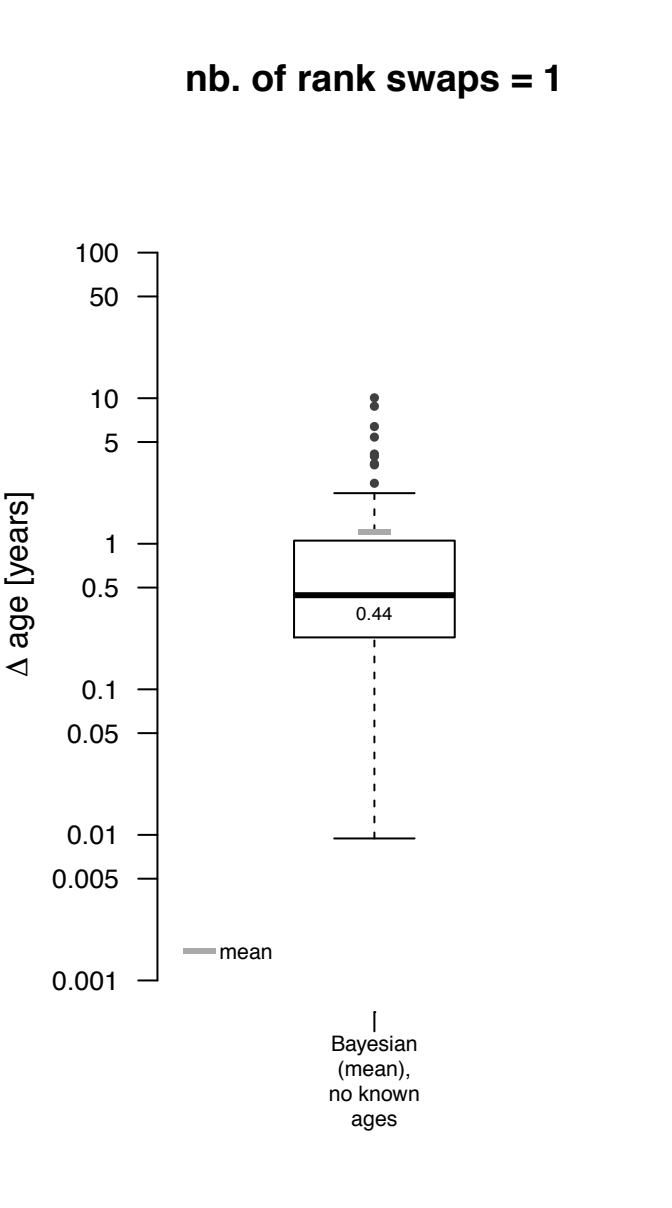

Avg. error on true age = 0.594
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.15
0.57
0.28
0.29
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages


Avg. error on true age = 0.84
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.4
0.66
0.31
0.32
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

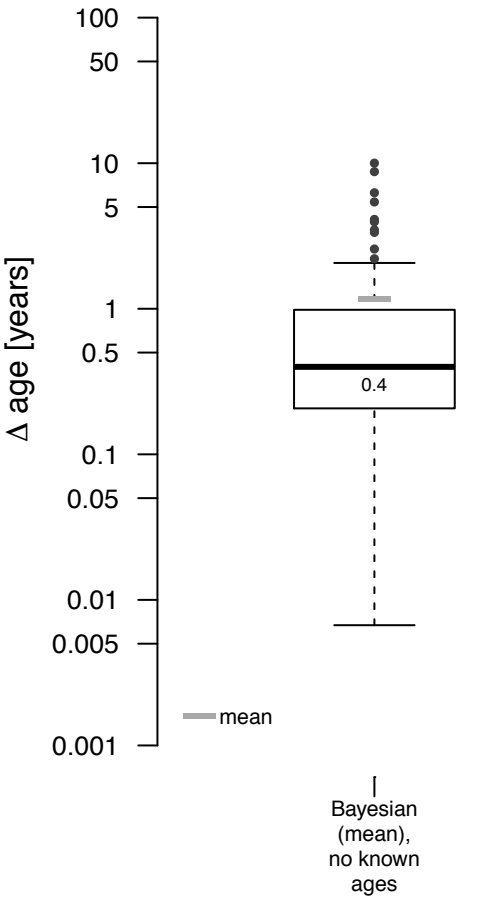

Avg. error on true age = 1.026
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.87
0.72
0.26
0.31
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages

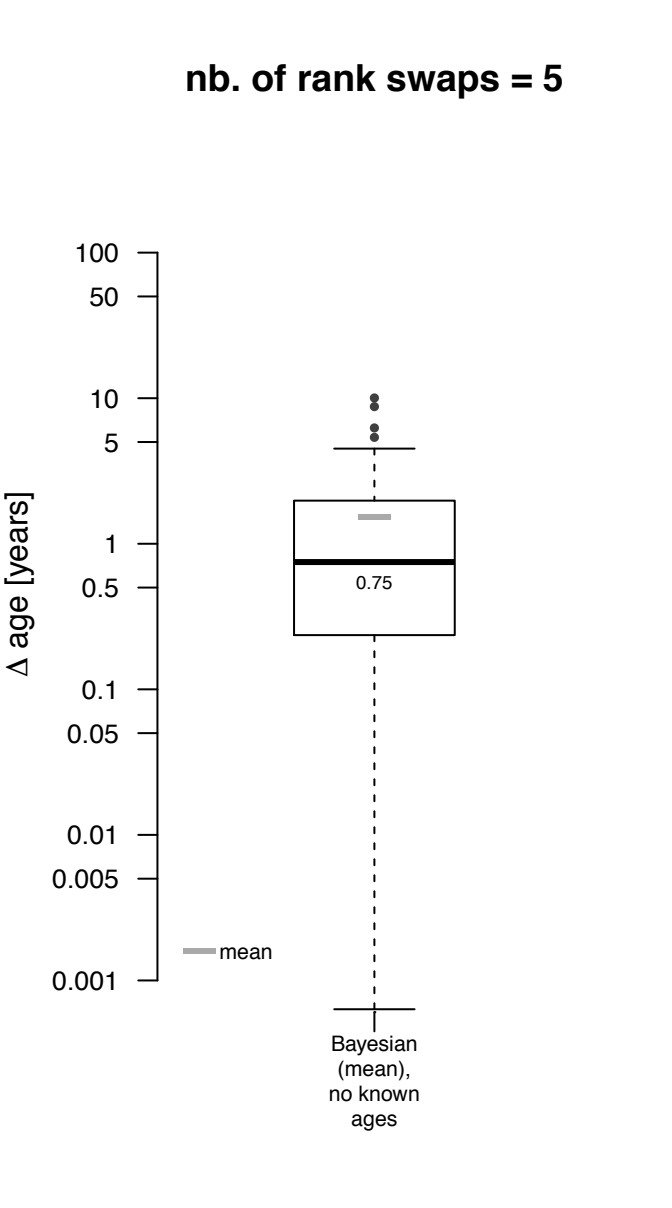

Avg. error on true age = 1.295
Δ age [years]
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
1.2
0.8
0.3
0.25
mean
5th−order polynomial
LOESS
Bayesian (mean)
Bayesian (mean), no known ages


Bayesian (mean), no known ages
Median error
0.5
1.0
1.5
2.0
0
10
20
30
40
50
60
70
% erroneous ranks

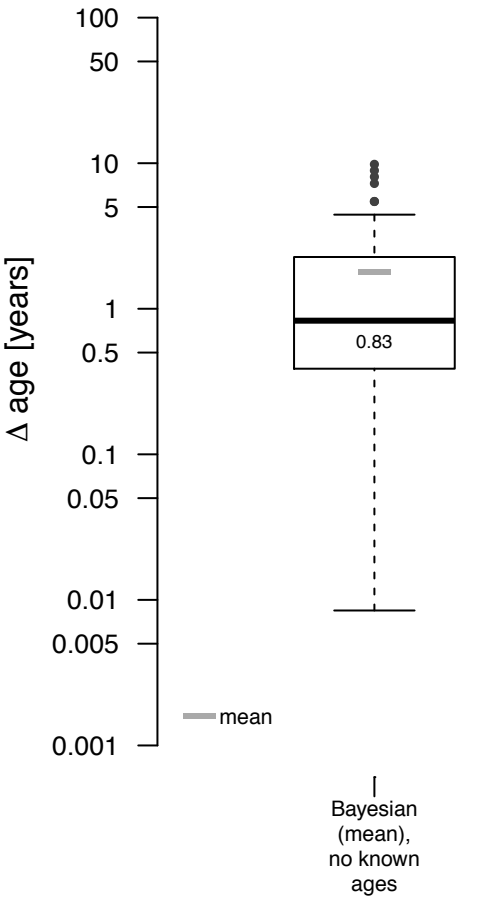

nb. of rank swaps = 1
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
0.44
mean
Bayesian
(mean),
no known
ages

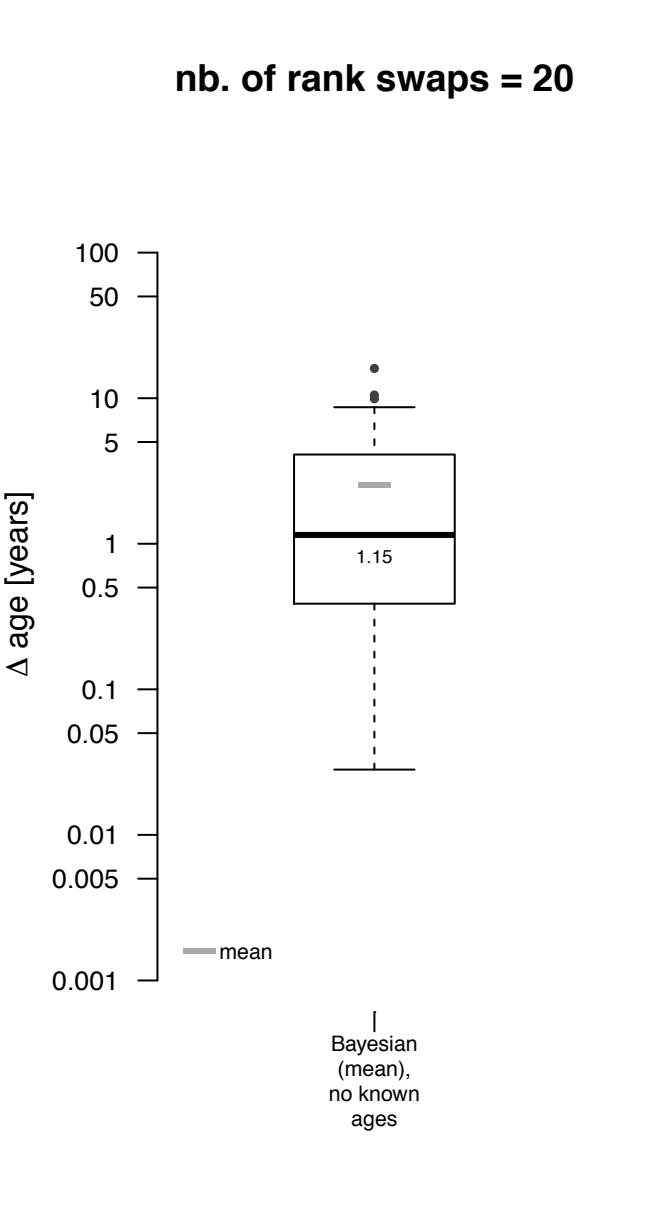
nb. of rank swaps = 2

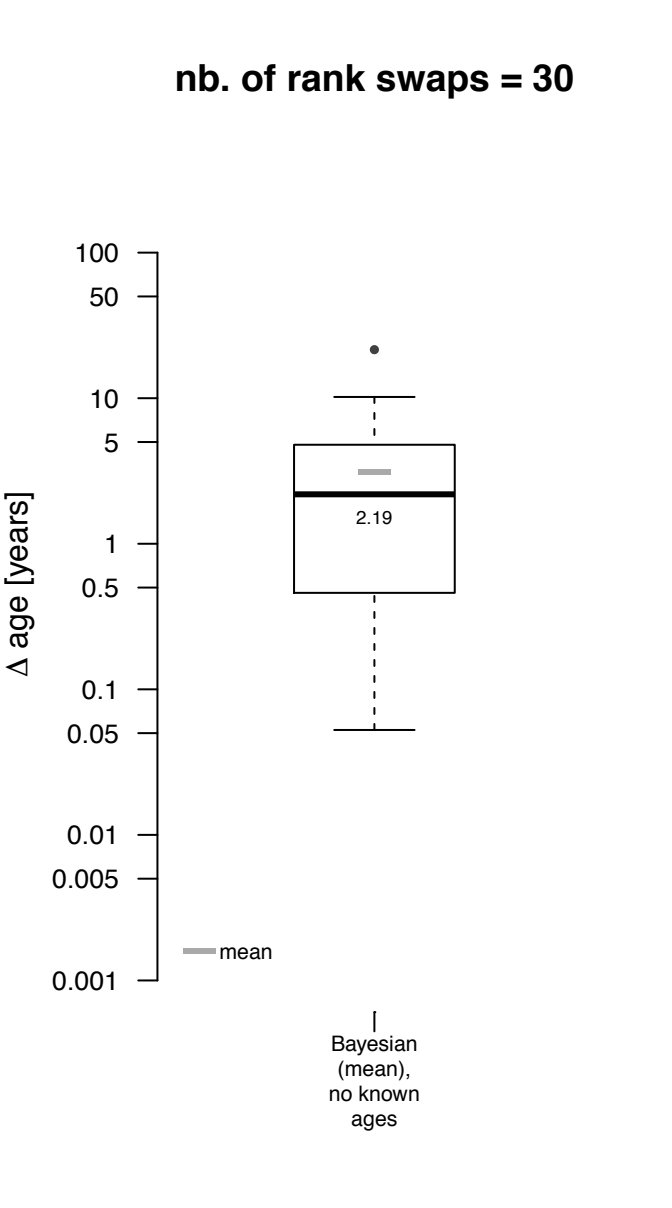



nb. of rank swaps = 5
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
0.75
mean
Bayesian
(mean),
no known
ages

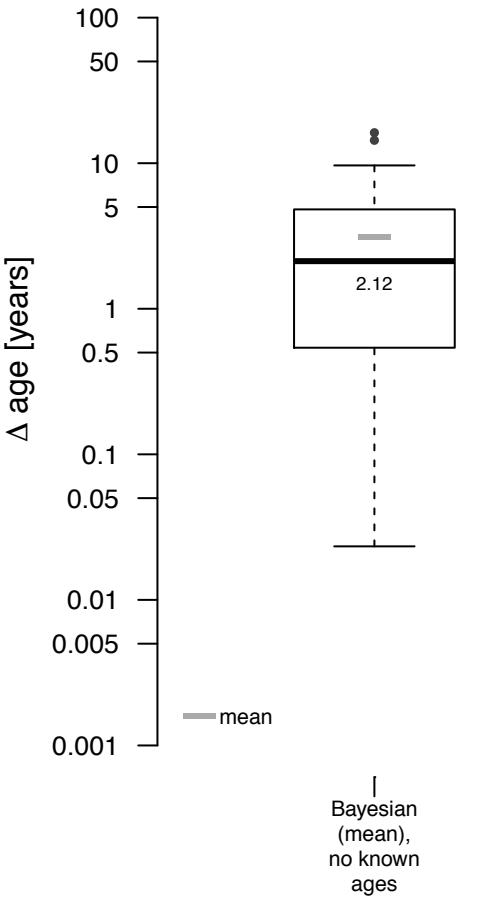
nb. of rank swaps = 10

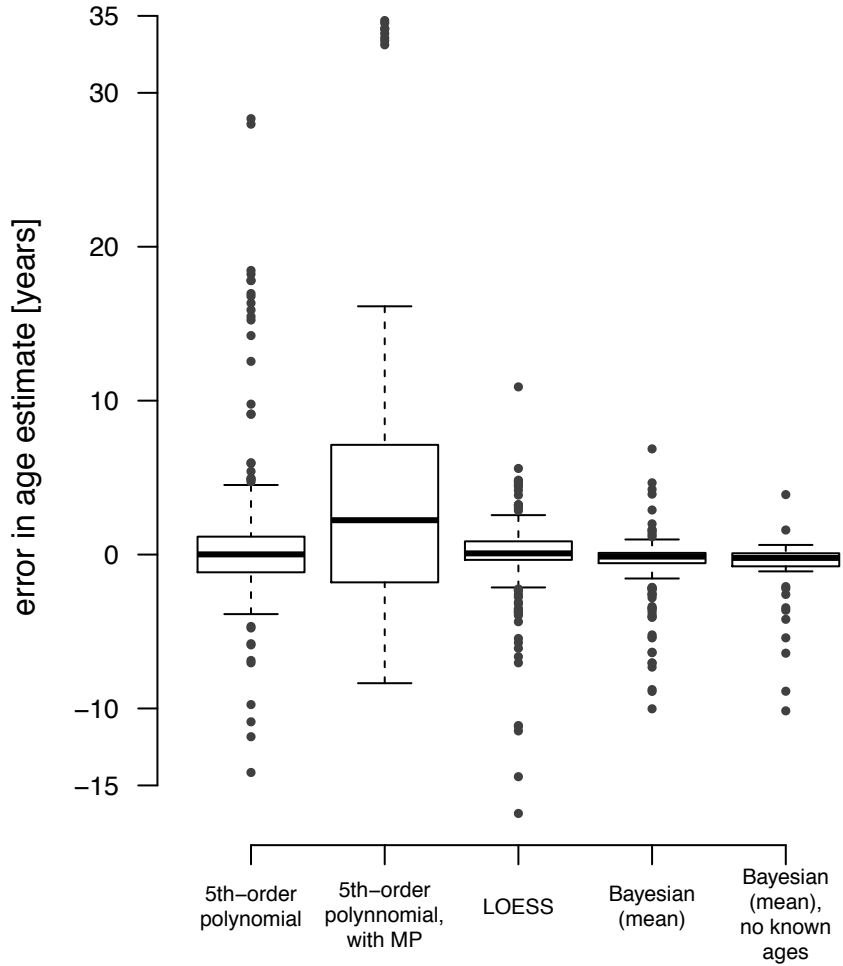



nb. of rank swaps = 20
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
1.15
mean
Bayesian
(mean),
no known
ages


nb. of rank swaps = 30
100
50
10
5
1
0.5
0.1
0.05
0.01
0.005
0.001
Δ age [years]
2.19
mean
Bayesian
(mean),
no known
ages

nb. of rank swaps = 50




error in age estimate [years]
35
30
20
10
0
−10
−15
5th−order polynomial
5th−order polynnomial, with MP
LOESS
Bayesian (mean)
Bayesian (mean), no known ages