

Predicting mortality from human faces

Dominika Dykiert, PhD^{1,2}, Timothy C. Bates, PhD^{1,2}, Alan J. Gow, PhD^{1,2}, Lars Penke, PhD^{1,2}, John M. Starr, MA, FRCPE^{1,3}, and Ian J. Deary, PhD, FRCPE^{1,2}.

¹ Centre for Cognitive Aging and Cognitive Epidemiology, University of Edinburgh

² Department of Psychology, University of Edinburgh

³ Geriatric Medicine Unit, University of Edinburgh

The manuscript contains 6129 words and 4 tables.

Correspondence:

Ian Deary (i.deary@ed.ac.uk) or Timothy Bates (tim.bates@ed.ac.uk)

Centre for Cognitive Ageing and Cognitive Epidemiology

Department of Psychology

University of Edinburgh

7 George Square

Edinburgh EH8 9JZ

United Kingdom

Tel. (+44) 131 650 3452

Abstract

Objective: To investigate whether and to what extent mortality is predictable from ratings of facial photographs of older people. **Methods:** High quality facial photographs of 292 members of the Lothian Birth Cohort 1921, taken at the age of about 83 years, were rated in terms of apparent age, health, attractiveness, facial symmetry, intelligence, and well-being by 12 young–adult raters. Cox proportional hazards regression was used to study the associations between these ratings and mortality over a 7-year follow-up period. **Results:** All ratings had adequate reliability. Concurrent validity was found for facial symmetry and intelligence (as determined by correlations with actual measures of fluctuating asymmetry and Raven’s Standard Progressive Matrices score, respectively), but not for the other traits. Age rated from facial photographs, adjusted for sex and chronological age, was a significant predictor of mortality (HR, 1.36; 95% CI, 1.12 to 1.65) and remained significant even after controlling for concurrent, objectively-measured health and cognitive ability, and the other ratings. Rated health, adjusted for sex and chronological age, significantly predicted mortality (HR, 0.81; 95% CI, 0.67 to 0.99) but not when rated age or objectively-measured health status were controlled. Rated attractiveness, symmetry, intelligence and well-being were not significantly associated with mortality risk. **Conclusions:** Rated age is a significant predictor of mortality risk among older people, with predictive value over and above that of objective or rated health status, and cognitive ability level.

Keywords: Rated age, face rating, mortality, survival

It is common to hear people comment of themselves or others that they look young (or old) for their age. Here, we ask whether apparent age, rated from facial photographs, predicts longevity. We also address this question for ratings of five additional factors—four believed to relate to longevity, namely: health (1), attractiveness (2), intelligence (3) and well-being (4), and symmetry, which has been linked to developmental stability, health, and age-related cognitive decline (5,6).

It has been reported that the age an individual reports himself of herself to feel reflects how successfully they are ageing. For example, among individuals aged between 65 and 84 years, those who felt physically or mentally older than their chronological age were twice as likely to die during a 13-year follow-up period than those who felt younger than their chronological age (7). Moreover, self-rated perceived age appears to be a stronger predictor of mortality than self-reported health status (8).

Age as perceived by others (which from now on will be referred to as *rated age*) has also been suggested as a biomarker of ageing (9). In the Baltimore Longitudinal Study of Aging (BLSA), 1,086 men aged between 17 and 102 had their chronological age rated by a physician, prior to a physical examination. These ratings were significantly associated with participants' scores on 7 out of 24 markers of 'biological age' (maximum breathing capacity, serum globulin, plasma glucose, creatinine clearance, two measures of tapping time, and foot reaction time).

A subsequent study on the BLSA sample showed that rated age was also related to mortality. With chronological age controlled, ratings of age performed at the first study visit were, on average, 1 year lower for the survivors than for the deceased (10). However, as already mentioned, the ratings of the BLSA participants' age were performed by physicians in the context of a medical examination. Although

the estimate was made at the beginning of an appointment, before any medical information was gathered, a face-to-face contact makes available a number of cues that may affect estimates of age. For example, the speed with which a participant enters the room, whether or not they use a walking stick, what they say in the initial interaction with the physician, their mood and perceived level of well-being may all affect the rated age. Therefore, it is not clear which characteristic or characteristics of an individual are reflected in their rated age, but it is probably more than their faces provide alone. The present study used ratings of age based solely on a facial photograph of each subject. Using static images of faces eliminates the cues mentioned above and allows a test of the hypothesis that the face (independent of verbal, mobility, and other cues) signals biological age.

Analyses in the Longitudinal Study of Aging Danish Twins (LSADT) demonstrated that rated age can be obtained from facial photographs with high inter-rater reliability and that these ratings are independent of the sex, age, and profession of the raters (11, 12). Rated age was significantly associated with mortality over a 7-year follow-up period even after controlling for sex, chronological age, grip strength, and cognitive and physical functioning, with a 3% increase in the risk of mortality per year increase in rated age (12). Because rated age was also associated with age-related cognitive and physical phenotypes, Christensen et al. (12) concluded that rated age obtained from static facial photographs is a robust biomarker of ageing.

The aim of the present study was to attempt to replicate and extend the findings of Christensen et al. (11, 12). The LSADT included individuals aged between 70 and 99 years; by contrast, our sample was a narrow age-cohort spanning only 2.6 years, with a mean age of 83.3. Therefore, the variation in the apparent age of the faces in the present study was almost entirely unrelated to differences in chronological

age, with a standard deviation for age of only 0.5 years. In addition to measuring rated age, we also included ratings of additional traits previously associated with cognitive decline, longevity or mortality. Among these were: self-rated health (1) and well-being (4), facial attractiveness rated by others (2), objectively-measured facial symmetry (6), and psychometrically-tested general intelligence (3). Including ratings of these additional traits allowed us to test whether they predict mortality, and whether the predictive power of rated age reflects cues of variation on some other traits, such as rated health or rated well-being.

METHODS

Participants

Participants were members of the Lothian Birth Cohort 1921 (LBC1921). The sample has been extensively described elsewhere (13, 14, 15) and is only outlined here with information relevant to the present study. LBC1921 includes individuals who were born in 1921, who took part in the Scottish Mental Survey of 1932 (16), and who lived in Edinburgh or its surroundings at the time of recruitment for the LBC1921 study, in 1999. These participants were invited to take part in three waves of the LBC1921 study. Wave 1 of testing started in 1999 when the participants were aged about 79 years and 565 individuals participated, providing at least partial data; wave 2 started in 2003 (age 83) with 413 individuals taking part; wave 3 started in 2007 (age 87) and 237 individuals were tested. The main analyses reported here were based on the sample of 292 individuals (133 males, 159 females), whose facial photographs from wave 2 were available.

Stimulus photographs

High resolution colour photographs of participants' faces (with neutral expression) were taken at wave 2, when the participants were aged between 82.0 and 84.6 years (Mean = 83.3, SD = 0.5). All photographs were taken under the same lighting, with the same digital camera and from the same distance.

The photographs were presented one at a time on a 22-inch high-quality cathode ray tube computer monitor screen. The question to be answered was presented immediately below the photo, with a sliding scale device used by the rater to enter ratings. The questions and their corresponding response scales are outlined in Table 1. Except for age, the default cursor position was in the middle of the respective rating scale (the rating for age was anchored at the low end of the scale).

Procedure

Photograph ratings

Twelve university student raters (6 males, 6 females) rated all photographs for age, health, attractiveness as a young person, facial symmetry, intelligence, and well-being. The questions were answered in the same order by all raters, and ratings were acquired for all subjects before proceeding to the next question; that is, all raters first assessed the age of all participants, followed by their health, attractiveness, and so forth. The ratings obtained from the 12 raters were averaged for each question to obtain a single measure of rated age, health, attractiveness, symmetry, intelligence, and well-being for each individual.

Insert Table 1 about here

Actual measures

At each wave of testing, a number of cognitive and physical measures, as well as various demographic and lifestyle information, were obtained from the LBC1921 participants. Measures used in the present study, which were mostly collected at wave 2, are described in detail below.

Health measures. Grip strength and blood pressure (BP) were measured as part of wave 2 testing. Grip strength was measured using a dynamometer with participants in a sitting position. Participants had their upper arm positioned vertically and their forearm at a right angle to the upper arm. Three measurements were taken from each hand, and the greatest measurement obtained from the dominant hand was used. Sitting systolic and diastolic BP were obtained manually using a cuff and sphygmomanometer. Self-reported information about the history of disease and use of medication were obtained during an interview. History of disease was coded as 1 if a participant had a history of any of the following: cardiovascular disease, cerebrovascular disease, neoplasia, hypertension, diabetes, or thyroid disease, and 0 otherwise; use of medication was coded 1 if a participant was using any medication at wave 2, and 0 otherwise.

Fluctuating asymmetry (FA). FA was calculated from the photographs taken at wave 2, according to the method proposed by Simmons et al. (17) and correcting for directional asymmetry (see reference 6, for a detailed description of the procedure). The data were transformed using a natural logarithm prior to the analyses and, for consistency with the ratings of symmetry, reversed so that higher values reflected greater symmetry.

Cognition. Raven's Standard Progressive Matrices (SPM; 18) and Mini Mental State Examination (MMSE; 19) were administered at wave 2. Raven's SPM is

a test of non-verbal reasoning, and the score (number of correct responses within a 20-minute time limit) was used here as a measure of general intelligence. MMSE is a 30-item untimed test, commonly-used as a screening tool for cognitive impairment, with a score below 24 often used to indicate possible dementia (20).

Well-being. No measure of well-being was available from wave 2. However, participants completed a 5-item Satisfaction With Life Scale (21) at waves 1 and 3, and these were used here. For consistency with the ratings of facial photographs, participant's agreement (on a 7-point Likert-type scale from strongly agree to strongly disagree) with the statement: "In many ways my life is close to my ideal" was selected for the present study as a measure of well-being.

Smoking. Participants provided information about their smoking habits at wave one. Self-reported smoking status (ever vs. never smoker) and number of cigarettes smoked per day were used in the present study.

Vital status and survival time

All LBC1921 participants are continuously followed up for mortality. At the beginning of the present investigation, the last recorded date of death was 18/01/2011. Survival time in days was calculated from the date on which a photograph was taken until death or, for participants who were alive, until the censoring date for this study (18/01/2011).

Statistical Analysis

The reliabilities of facial ratings of age, health, attractiveness, symmetry, intelligence and well-being were examined using the intra-class correlation (ICC). Because average ratings were used in the present study, and the variability due to

systematic differences between the raters was not considered relevant, we employed ICCs for the average ratings using consistency definition; that is, type (C, k) coefficient (22). This type of coefficient is equivalent to the commonly-used measure of internal consistency, Cronbach's alpha.

To assess the validity of the ratings, Pearson correlation coefficients were obtained for the association between each mean rating and an actual level of the relevant trait (an objective measure was chosen wherever possible). These were: for rated age—chronological age in days at the time the photograph was taken; for health—grip strength, disease history and use of medication; for ratings of symmetry—FA; for rated intelligence—Raven's SPM test score. Because no measures of subjective well-being were available from wave 2, measures obtained at waves 1 and 3 were used to test whether ratings based on photographs correlate with the actual level of well-being reported by the participants. There was no measure of actual attractiveness in youth available from the dataset; therefore, we did not test the validity of attractiveness ratings.

The principal analyses were concerned with the ability of the facial ratings to predict mortality. Cox proportional hazards regression was used to model the relationship of each rating with survival. First, simple models were fitted for each rating, with the facial rating as a predictor, while controlling for sex and the small amount of chronological age variation in the sample during the period in which the photographs were taken. Then, in order to establish the extent to which the contribution of the facial ratings to mortality risk was independent, models were re-run, controlling for additional risk factors. We controlled for BP (systolic and diastolic) and grip strength and, subsequently, for cognition (Raven's SPM score and MMSE at wave 2), in addition to age and gender. For the purpose of these analyses all

ratings were converted to standardised variables (mean = 0, SD = 1), so that the hazard ratios refer to a standard deviation difference in each rating.

RESULTS

Descriptive statistics, correlations and reliability estimates of the ratings are shown in Table 2. Reliabilities of the ratings of age, health, attractiveness, symmetry, intelligence and well-being, as assessed by ICCs, ranged from .70 to .85 (see Table 2). Rated age had highest reliability and rated intelligence the lowest. Despite the comprehensively-adequate reliability, not all ratings correlated significantly with the actual measures of their respective traits. Only symmetry ($r = .27$, $p < .001$) and intelligence ($r = .15$, $p = .009$) demonstrated such concurrent validity, with modest effect sizes.

Despite there being very little variance in chronological age of the participants (range 82.0 to 84.6 years) the rated age spanned over 20 years (63.5 to 85.3 years). Mean rated age in this study (74.2 years) was lower than the actual chronological age of the participants (83.3 years). There were no sex differences in rated health ($t(290) = 0.21$, $p = .833$), intelligence ($t(290) = 1.72$, $p = .087$), or well-being ($t(290) = -1.34$; $p = .182$). Females were rated as slightly older (mean = 74.71, SD = 3.92) than males (mean = 73.55, SD = 3.88); $t(290) = -2.54$, $p = .011$. There were also significant sex differences in rated attractiveness ($t(290) = -2.66$, $p = .008$) and symmetry ($t(290) = -2.44$, $p = .015$): female faces were rated as, on average, more attractive (female mean = 5.14, SD = 0.97; male mean = 4.84, SD = 0.89), and more symmetrical (female mean = 5.23, SD = 0.99; male mean = 4.95, SD = 0.93).

Rated age was strongly negatively correlated with ratings of health, and weakly with attractiveness, intelligence and well-being; the correlation between rated

age and symmetry was not statistically significant (see Table 2). Older-looking individuals were perceived to have poorer health, to have been less attractive as young adults, to have lower intelligence and lower levels of well-being. The correlations between ratings other than of age were all positive and moderate to strong in effect size; positive ratings tended to covary.

Insert Table 2 about here

Out of the 292 participants whose photographs were rated, 108 died between the wave 2 testing session (2003 to 2005) and 18/01/2011. Survival analyses, performed separately for each facial rating, revealed that rated age and rated health both significantly predicted mortality (see Table 3). Both effects were in the expected direction, with those rated as looking older and less healthy more likely to have died during the follow-up period. When controlling for gender and chronological age, the hazard of dying increased by 36% per standard deviation increase in rated age (HR, 1.36; 95% CI, 1.12 to 1.65; $p = .002$), which is equivalent to an 8% increase per year of rated age. This was only slightly attenuated after adjustment for health (BP and grip strength) and cognition (Raven's SPM score and MMSE). With gender and chronological age in the model, the hazard of dying decreased by 19% per standard deviation increase in rated health (HR, 0.81; 95% CI, 0.67 to 0.99; $p = .037$). This association was attenuated (and no longer significant) after controlling for objective health variables. Rated attractiveness, symmetry, intelligence, or well-being did not significantly predict mortality.

Insert Table 3 about here

In order to evaluate the relative predictive power of rated age and health, both were entered together in a model controlling for gender and chronological age. With rated age in the model, rated health was no longer a significant predictor of mortality (HR, 0.94; 95% CI, 0.75 to 1.18; $p = .572$), whereas rated age predicted mortality with a 31% increase in the hazard of dying per standard deviation increase in the rated age (HR, 1.31; 95% CI, 1.05 to 1.65; $p = .019$). Finally, when all ratings and all covariates were entered in a fully-adjusted model (see Table 4), the hazard of dying increased by 34% per standard deviation increase in rated age (HR, 1.34; 95% CI, 1.06 to 1.70; $p = .014$); no other ratings were significantly related to the risk of death. Among the covariates, MMSE was the strongest predictor of mortality (HR, 0.84; 95% CI, 0.76 to 0.93; $p = .001$) and the effect of sex approached the conventional significance level (HR, 0.59; 95% CI, 0.34 to 1.01; $p = .056$), pointing towards greater hazard of death among males than females. Given gender differences in life expectancy (23), this analysis was repeated separately for males and females and revealed that rated age was a significant predictor of mortality for females (HR, 1.51; 95% CI, 1.10 to 2.08; $p = .011$) but not males (HR, 1.17; 95% CI, 0.81 to 1.69; $p = .395$) (see Table 4).

Insert Table 4 about here

To explore the possibility that rated age is associated with mortality because both are related to harmful health behaviours, we tested whether controlling for

smoking status (ever vs. never smoker), or number of cigarettes smoked per day, would attenuate hazard ratios of rated age. Including smoking information from wave 1 did not attenuate the association between rated age and mortality risk (data not shown).

DISCUSSION

Rated age and health, obtained from static facial photographs of people in their early 80s, predicted mortality over a follow-up period of about 7 years. Individuals who appeared, to young adults, to be older and less healthy were more likely to die than those who appeared younger and healthier. Rated age predicted mortality even after controlling for sex, chronological age, objective measures of health status, and measured cognitive ability. These results support those of Christensen et al. (12) who also found rated age to be a robust predictor of mortality in their sample of older adults. Whereas Christensen et al.'s (12) sample spanned 29 years (70 to 99), in our study the age range was very narrow (2.6 years). The present study, therefore, indicates that considerable individual differences in rated age are found even in an age-homogenous sample (the lowest and highest average ratings of age were 63.5 and 85.3 in this sample of 82 to 84-year-olds). Stratifying the results by sex suggested that rated age was a significant predictor of mortality in females but not males. This finding may indicate that greater rated age is linked to higher mortality risk only in females. However, it may also be specific to our older sample in which, due to sex differences in survival rates, females were less selected.

Another novel contribution of the present study was the consideration of additional ratings obtained from facial photographs: those of health, attractiveness, symmetry, intelligence, and well-being. Rated health did predict mortality until actual

health status or rated age were controlled. Therefore, at least in statistical terms, rated age was a better predictor of mortality than rated health. This finding is consistent with earlier findings of greater predictive power of self-perceived age than self-reported health (8). Rated attractiveness, symmetry, intelligence, or well-being did not predict mortality in the 7 year follow-up period.

What do the ratings reflect?

As evidenced by high ICC ratings of reliability, rated age, health, attractiveness, symmetry, intelligence and well-being in older people could be reliably estimated from facial photographs. However, not all of these traits were related to the actual measures of traits of interest. Rated age was not significantly correlated with chronological age, though this is perhaps to be expected given the very narrow range of actual age in the present sample. More interestingly, rated health was unrelated to objectively measured grip strength or to self-reported health status. Only ratings of facial symmetry and intelligence were related to their objectively measured equivalents (measured FA and Raven's SPM score, respectively).

The lack of association between rated and actual measures of health may be due to inherent difficulties of obtaining objective measures that capture overall physical health. Self-reported disease history might have included diseases successfully treated in the past and, therefore, not be an accurate indicator of health status at the time the photograph was taken. Use of medication is probably a better index of current health, but effective medication is expected to alleviate negative symptoms, perhaps making the individuals appear healthier than they would otherwise.

Among the rated traits, most of the ratings were significantly correlated. This reflects previous literature where, for instance, symmetry ratings were moderately positively correlated with ratings of health (24), attractiveness and intelligence (25) and attractiveness was correlated with rated health (26, 27) and intelligence (25). The larger set of ratings used in the present study allows an overview of these relationships not visible in comparisons restricted to a single pair of variables. Despite a methodology chosen to minimise the influence of ratings for each trait on other ratings for a given face (by rating all faces on one trait, before moving on to the next trait) the ratings were clearly not independent. This suggests a common factor on which younger-looking people are rated as healthier, and as having more symmetrical faces and higher well-being. Alternatively, the traits on which the faces were rated may actually be related; that is, younger-looking individuals may actually be healthier, more attractive, etc. In our study, only some ratings reflected the actual traits, which points towards the former explanation. However, a comprehensive study with ratings and actual measures of matching traits would be useful to shed more light on the issue.

Mean rated age was almost 10 years lower than the chronological age in the present sample. This could be taken to suggest that the LBC1921 members look ‘young for their age,’ which would be consistent with their better physical and cognitive functioning relative to those who were not tested at wave 2 (14). However, it could also be an artefact of the age range presented to the raters (50 to 100 years). A previous study (28) reported that although “a slight preference for the middle age option” was noted, (p. 360), the midpoint of the age range used for ratings did not exert a large influence on the rated age estimates. However, in our study mean rated age (74.2) corresponded more closely with the midpoint of the rating scale (75 years)

than with the chronological age of the sample (83.3 years), suggesting that the rating age range presented may influence the mean ratings. Importantly, it did not appear to alter the individual differences, which were the focus here.

Potential mechanisms linking rated age with mortality

One explanation for the association between rated age and mortality risk is in terms of exposure to harmful factors; the face may be a canvas on which deleterious influences are recorded. In this sample, facial skin ageing has been found to be related to oxidative stress, a general health marker (29). One factor associated with both accelerated skin ageing (30) and mortality (31, 32), is smoking. Although not collected at wave 2, smoking information was available from wave 1. Adding smoking status (ever vs. never smoker) or number of cigarettes smoked per day to a Cox regression model did not attenuate hazard ratios of rated age. In other words, rated age predicted mortality independent of smoking, which suggests that it is more than just a record of negative health behaviours. However, it should be noted that the LBC1921 comprises an ageing elite. For example, those participants who took part in wave 2 testing had higher scores on physical and cognitive tests from wave 1 than those who did not return for the second wave of testing (14). Therefore, the LBC1921 may represent a group of individuals who are somewhat resilient to factors which affected their less successfully ageing counterparts, who consequently ceased to participate or died. Hence, the lack of support for the mediation of the association between rated age and mortality in the present sample may not be readily generalisable to the wider population. It would be informative to perform a similar analysis on a younger sample to establish whether some of the variance in rated age is

shared with harmful lifestyles and health behaviours, which lead to morbidity and mortality.

Another explanation for the association between rated age and mortality, in terms of common genetic influences on both, was suggested by Christensen et al. (12). When Christensen et al. stratified their twin sample by zygosity they found that, among dizygotic twins, the likelihood of the older looking twin dying first increased with an increase in the discrepancy in rated age between them. In other words, the more different in rated age the twins, the greater the likelihood of the older looking twin dying first. However, this was not true for monozygotic twins (i.e. with genetic factors controlled) and Christensen et al. (12) concluded that common genetic factors may be affecting both rated age and survival. For example, genes affecting the state of the cardiovascular tissue (and consequently, the risk of mortality from a myocardial infarction) may, at the same time, affect the condition of the skin (and so, the rated age).

The relationship between rated age and survival is likely to result from a mixture of genetic and environmental influences. Therefore both these avenues of research are worth pursuing in the future. Understanding the mechanisms linking rated age and mortality would help in understanding the process of ageing and inform the development of interventions to promote healthy ageing and longevity.

Strengths and limitations

Among the greatest strengths of this study is its sample with a narrow age range, which is preferred to age-heterogeneous cross-sectional samples in the study of ageing (33). All members of the LBC1921 were born in one year (1921) and they were at a similar age when their photographs were taken. Therefore, the large

confounding effect of chronological age was minimised. We demonstrated that the range of rated age (21.8 years) was over 8 times that of actual age (2.6 years), confirming that there are notable individual differences in the apparent rate of ageing.

The photographs used to obtain the ratings were taken at the same place, from a fixed distance and with the same camera. Therefore, they were more standardised than photographs available to Christensen et al. (11, 12), which were taken at participants' residences, and so, under different conditions. A further strength was the presence of a wide range of measured and rated covariates.

One limitation of the present study is that the surviving LBC1921 members are a select group. However, even in this select group, rated age predicted mortality. Had the sample been less restricted, especially in health status, the predictive value of rated age would be even greater than reported here.

Conclusions

Rated age, health, attractiveness, symmetry, intelligence and well-being can be reliably estimated from static facial photographs. However, whether ratings other than of facial symmetry and intelligence capture objective variance in the traits they are meant to measure, remains to be determined; they did not do so here. Rated age and health are significant predictors of mortality among older people; the former survives adjustment for the latter, but not vice versa. Age, as perceived from human faces, is a robust predictor of mortality, with predictive value over and above that of health status (objective or rated) and cognitive ability.

REFERENCES

1. DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality prediction with a single general self-rated health question. *J Gen Intern Med* 2006;21:267-275.
2. Henderson JJA, Anglin JM. Facial attractiveness predicts longevity. *Evol Hum Behav* 2003;24:351–356.
3. Calvin CM, Deary IJ, Fenton C, Roberts BA, Der G, Leckenby N, Batty DB. Intelligence in youth and all-cause-mortality: systematic review with meta-analysis. *Int J Epidemiol* 2010; *doi: 10.1093/ije/dyq190*.
4. Chida Y, Steptoe A. Positive psychological well-being and mortality: A quantitative review of prospective observational studies. *Psychosom Med* 2008;70:741-756.
5. Thornhill R, Moller AP. Developmental stability, disease and medicine. *Biol Rev Camb Philos Soc* 1997;72:497-548.
6. Penke L, Bates TC, Gow AJ, Pattie A, Starr JM, Jones BC, Perrett DI, Deary IJ. Symmetric faces are a sign of successful cognitive aging. *Evol Hum Behav* 2009;30:429–437.
7. Uotinen V, Rantanen T, Suutama T. Perceived age as a predictor of old age mortality: a 13-year prospective study. *Age Ageing* 2005;34:368–372.
8. Markides KS, Pappas C. Subjective age, health, and survivorship in old age. *Res Aging* 1982;4:87-96.
9. Borkan GA, Norris AH. Assessment of biological age using a profile of physical parameters. *J Gerontol* 1980;35:177-184.
10. Borkan GA, Bachman SS, Norris AH. Comparison of visually estimated age with physiologically predicted age as indicators of rates of aging. *Soc Sci Med* 1982;16:197-204.

11. Christensen K, Iachina M, Rexbye H, Tomassini C, Frederiksen H, McGue M, Vaupel JW. "Looking old for your age": Genetics and mortality. *Epidemiology* 2004;15:251-252.
12. Christensen K, Thinggaard M, McGue M, Rexbye H, Hjelmberg JvB, Aviv A, Gunn D, van der Ouderaa F, Vaupel JW. (2009). Perceived age as clinically useful biomarker of ageing: cohort study. *BMJ*, 339, b5262.
13. Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence on later life: following up the Scottish Mental Surveys of 1932 and 1947. *J Pers Soc Psychol* 2004; 86:130–47.
14. Gow AJ, Johnson W, Pattie A, Whiteman MC, Starr J, Deary IJ. Mental Ability in Childhood and Cognitive Aging. *Gerontology* 2008;54:177–186.
15. Gow AJ, Johnson W, Pattie A, Brett CE, Roberts B, Starr JM, Deary IJ. Stability and change in intelligence from age 11 to ages 70, 79, and 87: The Lothian Birth Cohorts of 1921 and 1936. *Psychol Aging* 2011;26:232-240.
16. Scottish Council for Research in Education. *The Intelligence of Scottish Children: A National Survey of an Age-Group*. London: University of London Press; 1933.
17. Simmons LW, Rhodes G, Peters M, Koehler N. Are human preferences for facial symmetry focused on signals of developmental instability? *Behav Ecol* 2004;15: 864–871.
18. Raven JC, Court JH, Raven J. *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: HK Lewis; 1977.
19. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state': A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189 –98.

20. Lezak MD, Howieson DB, Loring DW. *Neuropsychological Assessment*. 4th ed. Oxford: Oxford University Press, Inc; 2004.
21. Diener E, Emmons, RA, Larsen RJ, Griffin S. The satisfaction with life scale. *J Pers Assess* 1985;49:71-75.
22. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1:30-46.
23. Barford A, Dorling D, Davey Smith G, Shaw M. Life expectancy: women now on top everywhere. *BMJ* 2006;332:808; doi: 10.1136/bmj.332.7545.808.
24. Rhodes G, Zebrowitz LA, Clark A, Kalick M, Hightower A, McKay R. Do facial averageness and symmetry signal health? *Evol Hum Behav* 2001;22:31–46.
25. Zebrowitz LA, Hall JA, Murphy NA, Rhodes G. Looking smart and looking good: Facial cues to intelligence and their origins. *Pers Soc Psychol Bull* 2002;28:238-249.
26. Jones BC, Little AC, Penton-Voak IS, Tiddeman BP, Burt DM, Perrett DI. Facial symmetry and judgements of apparent health Support for a “good genes” explanation of the attractiveness–symmetry relationship. *Evol Hum Behav* 2001;22:417–429.
27. Kramer RSS, Ward R. Internal facial features are signals of personality and health. *Q J Exp Psychol* 2010;63:2273-2287.
28. Gunn DA, Murray PG, Tomlin CA, Rexbye H, Christensen K, Mayes AE. Perceived age as a biomarker of ageing: a clinical methodology. *Biogerontology* 2008;9:357-364.
29. Allerhand M, Ooi ET, Starr RJ, Alcorn M, Penke L, Drost E, MacNee W, Deary I J, Starr JM. Skin ageing and oxidative stress in a narrow-age cohort of older

- adults. *European Geriatric Medicine* 2011;2:140-144. doi:
10.1016/j.eurger.2011.04.009
30. Morita A. Tobacco smoke causes premature skin aging. *J Dermatol Sci*
2007;48:169-175.
31. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking; a
second report on the mortality of British doctors. *BMJ* 1956;2:1071-1081.
32. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50
years' observation on male British doctors. *BMJ* 2004;328:1519;
doi:10.1136/bmj.38142.554479.AE.
33. Hofer S, & Sliwinski M. Understanding Ageing: An evaluation of research
designs for assessing the interdependence of age-related changes. *Gerontology*
2001;47:341-352.

Table 1. Details of face photographs ratings

Aspect	Question	Rating Scale
Age	How old is this person?	50-100 (initialised at 50)
Health	How healthy is this person?	0-7 (Very Unhealthy - Very Healthy)
Attractiveness	How attractive was this person as a young adult?	0-10 (Very Unattractive - Very Attractive)
Symmetry	How symmetrical is this person's face?	0-10 (Very Asymmetrical - Very Symmetrical)
Intelligence	How intelligent is this person?	50-140 (Very low - Very High) ^a
Well-being	Shown the question "My life is close to ideal", this person would respond:	0-7 (Strongly Disagree - Strongly Agree)

^a An information message: "100 is average" was displayed on the screen for intelligence rating

Table 2. Inter-correlations between ratings and correlations with actual measures

	Rated age	Rated health	Rated attractiveness	Rated symmetry	Rated intelligence	Rated well-being	M	SD	Reliability	Correlations with the actual measures ^a
Rated age	-	-.55***	-.18**	-.04	-.14*	-.19***	74.18	3.93	.85	.05 (chronological age)
Rated health		-	.63***	.32***	.51***	.58***	3.38	0.76	.79	.10 (grip strength) .04 (history of disease) ^b -.03 (use of medication) ^b
Rated attractiveness			-	.47***	.60***	.63***	5.00	0.95	.82	-
Rated symmetry				-	.33***	.40***	5.10	0.97	.80	.30*** (FA)
Rated intelligence					-	.52***	102.03	5.41	.70	.15** (Raven's SPM)
Rated well-being						-	4.15	0.73	.80	.03 (well-being wave 1) .09 (well-being wave 3)

Table 2. [Continued]

Note. The correlations between ratings are based on $n = 292$; ns for correlations with actual measures vary from 202 to 292 due to missing data on some variables. FA = fluctuating asymmetry, SPM = Standard Progressive Matrices

^a Actual measures were as follows: age - chronological age in days at the time the photograph was taken; health – grip strength at wave 2, history of disease (cardiovascular, cerebrovascular, neoplasia, hypertension, diabetes, or thyroid disease), and use of medication; attractiveness - no actual measure was available from the dataset; symmetry – FA measured from facial photographs, reversed so that higher FA indicates more symmetry; intelligence – score obtained on Raven's SPM test at wave 2; well-being – self-report of well-being obtained at wave 1 (age 79) and wave 3 (age 87).

^b Point bi-serial correlation

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3. Hazards ratios and 95% confidence intervals for the association between each rating and death from models fitted separately for each rating.

Variable	Adjustments								
	Chronological age and sex			Chronological age, sex, BP and grip strength			Chronological age, sex, BP, grip strength, and cognition		
	HR	CI (95%)	p	HR	CI (95%)	p	HR	CI (95%)	p
Rated age	1.36	1.12 - 1.65	.002	1.33	1.10 - 1.61	.004	1.33	1.09 - 1.62	.004
Rated health	0.81	0.67 - 0.99	.037	0.83	0.68 - 1.01	.056	0.86	0.71 - 1.05	.149
Rated attractiveness	0.86	0.71 - 1.04	.125	0.86	0.71 - 1.04	.128	0.89	0.73 - 1.07	.216
Rated symmetry	0.85	0.70 - 1.02	.087	0.86	0.72 - 1.04	.119	0.92	0.76 - 1.11	.392
Rated intelligence	0.94	0.78 - 1.15	.558	0.97	0.79 - 1.18	.726	1.00	0.82 - 1.22	.984
Rated well-being	0.93	0.77 - 1.13	.486	0.94	0.78 - 1.14	.543	1.00	0.82 - 1.21	.964

Note. Models were fitted separately for each rating.

CI = confidence interval, BP = blood pressure (sitting systolic and diastolic). Adjustment for cognition included Raven's Standard Progressive Matrices score and Mini Mental State Examination

Table 4. Hazard ratios and 95% confidence intervals for the association with death from models including all ratings and covariates

Variable	Whole sample (n = 292)			Males only (n = 133)			Females only (n = 159)		
	HR	CI (95%)	p	HR	CI (95%)	p	HR	CI (95%)	p
Rated age	1.34	1.06 - 1.70	.014	1.17	0.81 - 1.69	.395	1.51	1.10 - 2.08	.011
Rated health	1.04	0.75 - 1.45	.792	0.98	0.63 - 1.53	.940	1.19	0.73 - 1.94	.494
Rated attractiveness	0.85	0.63 - 1.14	.267	0.81	0.54 - 1.22	.315	0.86	0.55 - 1.35	.517
Rated symmetry	0.93	0.75 - 1.17	.551	0.87	0.65 - 1.15	.325	1.10	0.77 - 1.58	.587
Rated intelligence	1.10	0.85 - 1.42	.486	1.10	0.76 - 1.60	.600	1.09	0.74 - 1.59	.662
Rated well-being	1.10	0.84 - 1.43	.495	1.35	0.92 - 1.99	.127	0.86	0.56 - 1.30	.466
Gender	0.59	0.34 - 1.01	.056	-	-	-	-	-	-
Chronological age	1.00	1.00 - 1.00	.729	1.00	1.00 - 1.00	.184	1.00	1.00 - 1.00	.326
Systolic BP	1.00	0.99 - 1.01	.557	1.00	0.98 - 1.02	.928	0.99	0.98 - 1.01	.461
Diastolic BP	1.00	0.98 - 1.02	.890	0.99	0.96 - 1.02	.505	1.00	0.96 - 1.03	.922
Grip strength	0.97	0.94 - 1.00	.086	0.96	0.92 - 1.00	.045	1.01	0.94 - 1.08	.879
Raven's SPM score	0.99	0.97 - 1.02	.665	1.01	0.97 - 1.05	.587	0.98	0.94 - 1.01	.207
MMSE	0.84	0.76 - 0.93	.001	0.83	0.70 - 0.98	.026	0.85	0.73 - 0.99	.038

Note. CI = confidence interval, BP = blood pressure (sitting systolic and diastolic), SPM = Standard Progressive Matrices, MMSE = Mini Mental State Examination

