Title

When does brain aging accelerate? Dangers of quadratic fits in cross-sectional studies.

Journal

NeuroImage, 50(4)

ISSN

1053-8119

Authors

Fjell, Anders M
Walhovd, Kristine B
Westlye, Lars T
et al.

Publication Date

2010-05-01

DOI

10.1016/j.neuroimage.2010.01.061

Peer reviewed

NeuroImage 50 (2010) 1376–1383

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

When does brain aging accelerate? Dangers of quadratic fits in cross-sectional studies

Anders. M. Fjell [a,b,*,1], Kristine B. Walhovd [a,b,1], Lars T. Westlye [a], Ylva Østby [a], Christian K. Tamnes [a], Terry L. Jernigan [c,e,g,h], Anthony Gamst [c,f,i], Anders M. Dale [d,e,f]

[a] Center for the Study of Human Cognition, Department of Psychology, University of Oslo, Pb. 1094 Blindern, 0317 Oslo, Norway
[b] Department of Neuropsychology, Ullevaal University Hospital, Oslo, Norway
[c] Laboratory of Cognitive Imaging, University of California, San Diego and the VASDHS, 9500 Gilman Drive, La Jolla, CA 92093-0949, USA
[d] Multimodal Imaging Laboratory, University of California, San Diego, CA, USA
[e] Department of Radiology, University of California, San Diego, CA, USA
[f] Department of Neurosciences, University of California, San Diego, CA, USA
[g] Department of Cognitive Science, University of California, San Diego, CA, USA
[h] Department of Psychiatry, University of California, San Diego, CA, USA
[i] Biostatistics and Bioinformatics, University of California, San Diego, CA, USA

ARTICLE INFO

Article history:
Received 11 June 2009
Revised 16 December 2009
Accepted 19 January 2010
Available online 25 January 2010

Keywords:
Aging
Magnetic resonance imaging
Hippocampus
Quadratic functions
Polynomial models
Local smoothing
Smoothing spline
Nonparametric
FreeSurfer

ABSTRACT

Many brain structures show a complex, non-linear pattern of maturation and age-related change. Often, quadratic models ($\beta_0+\beta_1 age+\beta_2 age^2+\varepsilon$) are used to describe such relationships. Here, we demonstrate that the fitting of quadratic models is substantially affected by seemingly irrelevant factors, such as the age-range sampled. Hippocampal volume was measured in 434 healthy participants between 8 and 85 years of age, and quadratic models were fit to subsets of the sample with different age-ranges. It was found that as the bottom of the age-range increased, the age at which volumes appeared to peak was moved upwards and the estimated decline in the last part of the age-span became larger. Thus, whether children were included or not affected the estimated decline between 60 and 85 years. We conclude that caution should be exerted in inferring age-trajectories from global fit models, e.g. the quadratic model. A nonparametric local smoothing technique (the smoothing spline) was found to be more robust to the effects of different starting ages. The results were replicated in an independent sample of 309 participants.

Introduction

Over the past few years, research has demonstrated that most brain structures undergo a complex pattern of maturation and age-related change. For instance, the hippocampus shows a marked non-linear pattern of change throughout the lifespan (Allen et al., 2005; Jernigan and Gamst, 2005; Kennedy et al., 2008; Walhovd et al., 2005). Very often, non-linearity of age relationships is tested using quadratic or other polynomial models. A quadratic term is added to the list of predictors in a regression analysis, yielding a higher order polynomial function. If the quadratic term is significant, the brain structure in question can be said to have a non-linear age-trajectory (Allen et al., 2005; Good et al., 2001; Jernigan and Gamst, 2005; Kennedy et al., 2008; Lupien et al., 2007; Sowell et al., 2003; Sullivan et al., 1995; Terribilli et al., 2009; Walhovd et al., 2005). In addition, the trajectory of the curve may be used to describe the relationship between age and the brain structure, e.g. to determine when the hippocampus reaches its maximum volume, or how steep the subsequent decline is. This, however, may be problematic: First, to say that a relationship is non-linear is not the same as saying that it is quadratic. This is an example of a specification effect. Second, if the same quadratic model is fit to different sets of data, one will get different results. For example, the observed peaks of quadratic functions will inherently depend on the age range sampled. This can lead to completely erroneous inferences about features of the trend, for example the location of peaks. This is a localization effect. The aim of this report is to demonstrate biases associated with quadratic model fits, and hint at possible solutions.

The quadratic function is always a parabola, and the basic shape only differs in curvature and direction (whether it has a peak or a dip). Thus, it will not yield an accurate description of age-trajectories

* Corresponding author. Center for the Study of Human Cognition, Department of Psychology, University of Oslo, Pb. 1094 Blindern, 0317 Oslo, Norway. Fax: +47 22 84 50 01.
E-mail address: andersmf@psykologi.uio.no (A.M. Fjell).
[1] The authors have contributed equally to the paper.

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.01.061

characterized by steep increase in development, slow decline during adulthood, and then a sharper decline in older age. Further, the regression line depends on all data points, thus representing a global fit. This means that adding more data points to one part of the sample, e.g. including more children, will change the fit in distant parts of the age-range. Thus, estimation of age-decline after 60 years will ultimately depend on how early in the life-span sampling begins. Some aging studies sample from childhood (Courchesne et al., 2000; Sowell et al., 2007), some from young adulthood (Allen et al., 2005; Fjell et al., 2009; Raz et al., 2004; Walhovd et al., 2005), and some from middle-age or higher (Du et al., 2006; Greenberg et al., 2008; Van Petten, 2004). This variation may exert substantial effects on the observed age-trajectories, but has not been tested with real neuroanatomical data. In the present paper we demonstrate that when quadratic functions are used, the age at which one starts to sample has a systematic effect on the slope of the curve at all subsequent ages, fundamentally changing the interpretation of the age changes. Age-functions were tested for the volume of hippocampus in a large sample of 434 participants ranging from 8 to 85 years of age. The results were compared to the outputs of a nonparametric local smoothing model, the smoothing spline.

Materials and methods

Sample

The main sample was drawn from two ongoing longitudinal research projects at the Center for the Study of Human Cognition, Department of Psychology, at the University of Oslo (Neurocognitive Development/Cognition and Plasticity through the Life-Span). Procedures are presented in detail elsewhere (Fjell et al., 2008; Westlye et al., 2009). 434 right handed native Norwegian speakers without injury or disease known to affect CNS function participated (age 8–85 years, mean 41.4, SD = 22.0, 238 women/196 men). All MR scans were deemed free of significant injuries or conditions by a specialist in neuroradiology.

To assess the stability of the results, the main analyses were repeated in an independent sample of participants drawn from the Open Access Series of Imaging Studies (OASIS) database (http://www.oasis-brains.org/), a publicly available resource of cross-sectional MR scans. 316 participants from this database are non-demented. Of these, seven were rejected due to less than optimal scan quality, bringing the n down to 309 (age 18–94, mean 44.6, SD = 23.6, 194 women/115 men) . The details of this database are described in (Marcus et al., 2007). The youngest participants (<60 years) were interviewed about medical history and use of psychoactive drugs, while elderly participants (>60 years) underwent the Washington University Alzheimer Disease Research Center's full clinical assessment, excluding participants with dementia (clinical dementia rating >0) (Berg, 1984, 1988), active neurological or psychiatric illness, serious head injury, history of clinically meaningful stroke, use of psychoactive drugs, or gross anatomical abnormalities evident from MRI images. In addition, MMSE was assessed for all the elderly participants, with a range of 25–30.

MR acquisition and analysis

For the main sample, two repeated 3D T1-weighted Magnetization Prepared Rapid Gradient Echo (MP-RAGE) were acquired using a 12-channel head coil on the same 1.5-Tesla Siemens Avanto scanner (Siemens Medical Solutions, Erlangen, Germany), with the following parameters: TR/TE/TI/FA = 2400 ms/3.61 ms/1000 ms/8°, matrix 192 × 192, field of view = 192. Each volume consisted of 160 sagittal slices with voxel sizes 1.25 × 1.25 × 1.2 mm. The two MP-RAGEs were averaged during post-processing to increase the signal-to-noise-ratio (SNR). For the replication sample (OASIS), three to four 3D T1-weighted MP-RAGEs were acquired on the same 1.5-Tesla Siemens Vision scanner, with the following parameters: TR/TE/TI/FA = 9.7 ms/4.0 ms/20 ms/10°, matrix 256 × 256. Each volume consisted of 128 sagittal slices with voxel sizes 1.25 × 1.0 × 1.0 mm. The three to four SPGRs were averaged to increase SNR.

All datasets were processed and analyzed with FreeSurfer 4.01 (http://surfer.nmr.mgh.harvard.edu/) at the Neuroimaging Analysis Lab, Center for the Study of Human Cognition, University of Oslo, with the additional use of computing resources from the ~4000 CPUs titan grid cluster (http://hpc.uio.no) run by the Research Computing Services Group at USIT, University of Oslo. The automated procedure for volumetric measures of the different brain structures is described in detail by Fischl et al. (2002, 2004). A label was automatically assigned to each voxel in the MRI volume based on probabilistic information automatically estimated from a manually labelled training set. The training set included both healthy persons in the age range 18–87 years and a group of AD patients in the age range 60–87 years, and the classification technique employed a registration procedure that is robust to anatomical variability, including the ventricular enlargement typically associated with aging. The technique has been shown to be comparable in accuracy to manual labelling (Fischl et al., 2002; Fischl et al., 2004), and we have used it in previous studies of development (Walhovd et al., 2007) and aging (Walhovd et al., 2005, in press). Further, as the gross morphology of 8-year-old brains is rather similar to the adult brains, the procedure works well for children of this age. We also applied a newly developed atlas normalization procedure (Han and Fischl, 2007). Hippocampus was chosen as region of interest (ROI). The segmentation of the hippocampal formation included dentate gyrus, CA fields, subiculum/parasubiculum and the fimbria (Makris et al., 1999). Estimated intracranial volume (ICV) was used to correct the volumetric data. This was calculated by use of an atlas-based normalization procedure, where the atlas scaling factor is used as a proxy for ICV, shown to correlate highly with manually derived ICV ($r = 0.93$) (Buckner et al., 2004). An example of the segmentation for an individual participant from the main sample is shown in Fig. 1.

Statistical analyses

Quadratic regression models of the form residual hippocampal volume $= \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \varepsilon$ (sex and ICV regressed out) were tested for the age-ranges 8–85, 20–85, 30–85, 40–85, 50–85 and 60–85 years. The results were compared to a nonparametric local smoothing model, the smoothing spline, implemented in matlab. Given a sequence of data (X_i, Y_i); $i = 1, \ldots, n$, with $E(Y_i \mid X_i) = g(x_i)$, the smoothing spline estimate of g minimizes

$$\sum_{i=1}^{n} (Y_i - g(X_i))^2 + \lambda \int (g''(x))^2 dx,$$

where the smoothing parameter λ controls the trade-off between fidelity (closeness of $g(X_i)$ to Y_i) and smoothness (the size of the average second derivative of g). With no smoothing ($\lambda = 0$), g simply interpolates the data, whereas infinite smoothing ($\lambda = \infty$), g corresponds to the line fit by ordinary least-squares. We used an algorithm that optimizes smoothing level based on a version of Akaike's Information Criterion (AIC), i.e. the smoothing level that minimizes AIC for each analysis was chosen. AIC offers a relative measure of amount of information lost when a model is used to describe a set of data, and can be said to describe the trade off between bias and variance in the construction of statistical models. AIC rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. Thus, AIC attempts to find the model that best explains the data with a minimum of free parameters, in this case, with a largest possible smoothing level. With no smoothing, the smoothing spline will yield an extremely good

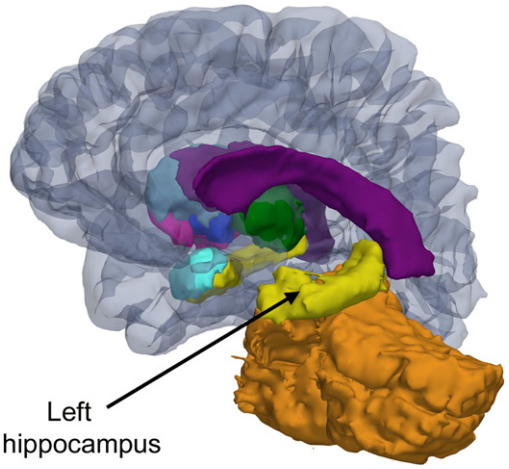

Fig. 1. Three-dimension rendering of hippocampus for a representative participant.

apparent fit to the data, but the model would be predictively inaccurate. AIC takes this into account by penalizing for degrees of freedom. For this analysis, yhat = Hy, where **H** is the smoothing matrix, the effective degrees of freedom of the fitted model is given by the trace of the matrix **H**, and an argument leading AIC produces the corresponding estimate of the model risk. Use of AIC to determine level of smoothing provides a way of obviating the need for arbitrarily chosen smoothing levels. The resulting AICs obtained at different smoothing levels for each of the age-ranges tested are shown in Fig. 2. For the analysis restricted to the sample above 60 years, however, the smoothing level that minimized AIC seemed too low based on visual inspection of the fit, so in the graph smoothing level was set to λ (lam) = $3 \times e^{-4}$ for this analysis. We also calculated the age where the expression $-\frac{d^2 f(age)}{d\, age^2}$ was largest, i.e. the point where the slope of the local smoothing curve changed the most (the second derivative). For the quadratic model, the second derivative is assumed to be constant across the life span, and hence the point of maximum acceleration cannot be determined.

While it is possible to test for non-linearity and to reject quadratic models, e.g. by use of adaptive tests (Fan et al., 2001), this was not the focus of the present paper. Rather, the local smoothing model was included to present an alternative to the global quadratic model. Still, to give the reader an opportunity to compare goodness of fit between the models, AIC was also calculated for the quadratic models. It is important to note that as AIC contains scaling constants, the absolute AIC values for the different models have no meaning. This implies that in the present paper, AIC can be used to compare smoothing splines with quadratic models, but not to compare how well the models perform across different age-ranges. To ease comparison of AIC between quadratic and smoothing spline models, we used Δ_I, which is the difference between AIC for the model and the lowest AIC (in this case, the difference between

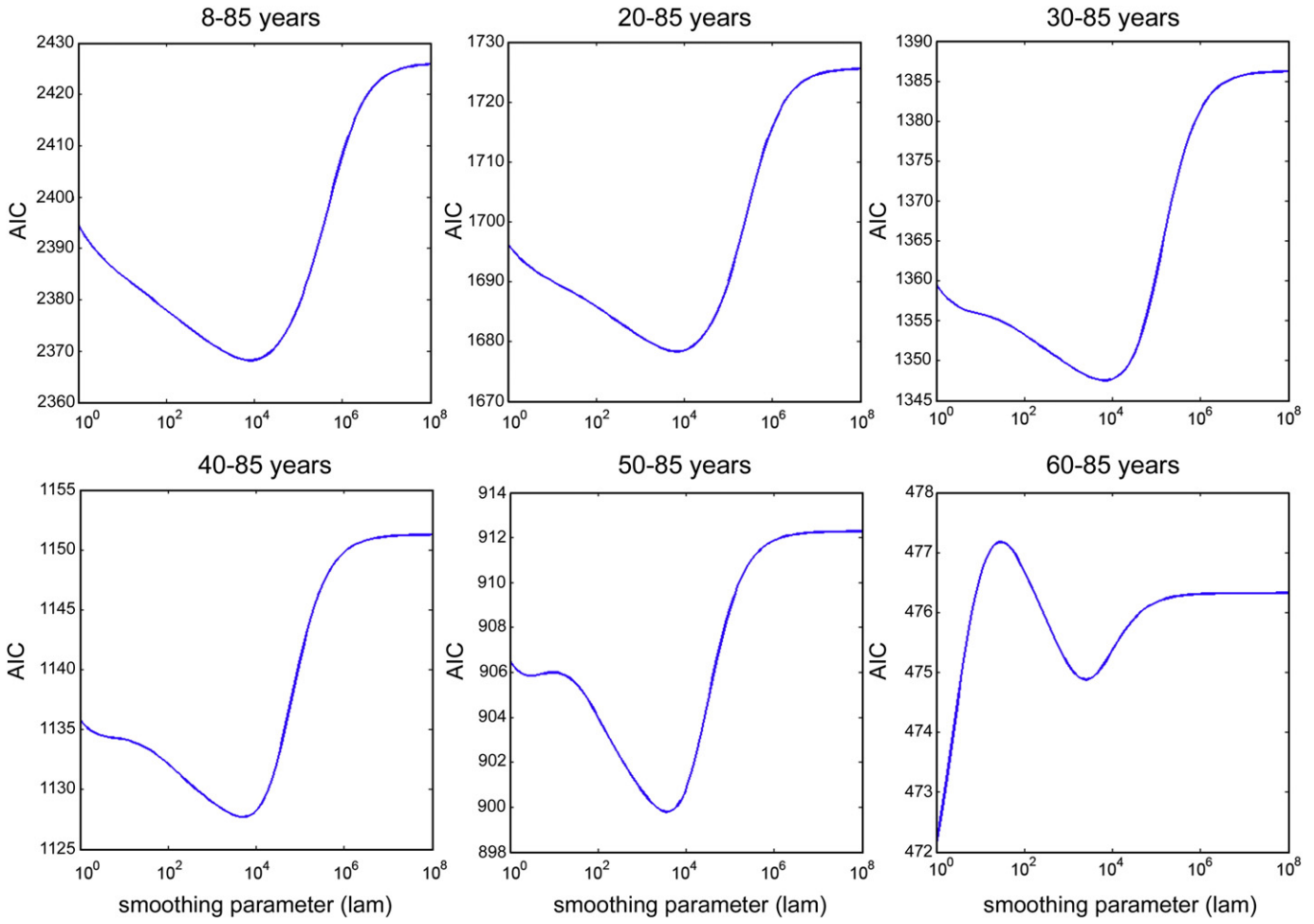

Fig. 2. Aikake's Information Criterion (AIC) as a function of smoothing level. For the smoothing spline model, AIC was calculated as a function of smoothing level. Apparent goodness of fit will increase with lower levels of smoothing, and so AIC takes degrees of freedom into account. Note that a low AIC is desirable, but that the absolute value of AIC gives less meaning, and that AIC cannot be used to compare models between the age-ranges tested. The figure shows that the optimal level of smoothing is almost identical for all age-ranges except 60–85 years, where the lowest AIC is obtained by no smoothing at all.

Table 1
Multiple quadratic regressions for different age spans in the main sample.

Hippocampal volume						
Age-span	Equation	*df*	R^2	*F*	*p* of age^2 ≤	*p* of *F* ≤
08–85 years	$0.346+0.025\times age-0.001\times age^2$	2, 431	0.44	165.81	2.5×10^{-11}	3.9×10^{-54}
20–85 years	$-0.228+0.049\times age-0.001\times age^2$	2, 327	0.43	123.14	1.3×10^{-9}	1.4×10^{-40}
30–85 years	$-2.293+0.121\times age-0.001\times age^2$	2, 266	0.41	93.07	3.1×10^{-10}	2.3×10^{-31}
40–85 years	$-4.685+0.198\times age-0.002\times age^2$	2, 227	0.41	79.29	2.0×10^{-7}	7.7×10^{-27}
50–85 years	$-6.580+0.254\times age-0.002\times age^2$	2, 189	0.43	70.36	8.6×10^{-5}	1.5×10^{-22}
60–85 years	$-9.370+0.333\times age-0.003\times age^2$	2, 111	0.44	43.85	0.051	9.2×10^{-15}

"*p* of age^2" is the *p*-value for the quadratic term when age is included as a simultaneous regressor, while "*p* of *F*" is the *p*-value for the expression.

the best model and the other model). As a rule of thumb, $\Delta_I<2$ would indicate that the two models are essentially indistinguishable with regard to goodness of fit, $\Delta_I>4$ would indicate considerable differences between the models, and $\Delta_I>10$ would indicate that the model has essentially no support. In addition to AIC, we also chose to run the smoothing spline with smoothing levels chosen to minimize the Bayesian information criterion (BIC), and compared the resulting BICs. Similar to AIC, BIC resolves the problem of overfitting by introducing a penalty term for the number of parameters in the model, but the penalty for additional parameters is stronger than that of the AIC, biasing it towards more smoothing and giving more consistent smoothing values.

We expected that the quadratic models would be more affected by changing the age-range of the sample analyzed than the smoothing spline models. To assess whether this would hold for more than one sample, the most relevant analyses were repeated in an independent sample of participants (referred to as the replication sample).

Results

Effects of age-range on the quadratic model

The full model as well as the quadratic term was highly significant for all age-ranges except when sampling began at 60 years, with inverted U-shaped curves (Table 1). When sampling began at 60 years, the quadratic terms was only marginally significant ($p=0.051$). Whenever sampling was skewed towards a higher starting point, the age at which volumes peaked was moved upwards and the estimated decline in the last part of the age-span became larger (Table 2 and Fig. 3). Peak estimated volume was 20 years when the models was tested with the full sample (8–85 years), increased to 29 when the sample was restricted to the participants from 20 years, and further increased to 41, 48 and 52 years when the youngest participants were 30, 40 and 50 years old, respectively. Further, the estimated age decline from 60 to 85 years increased whenever sampling was skewed towards a higher staring point, gradually from −1.64 *z*-scores when all participants were included to −2.62 *z*-scores when only participants from 60 years were included.

Effects of age-range on the smoothing spline model

The curves for the smoothing spline were highly similar regardless of at what age sampling began, with monotone decline for all sampling ranges. For instance, the point where the rate of change of the slope was highest (the second derivative) was 49 years regardless of whether sampling started at 8, 20 or 30 years, and was 47 for the age-range 40–85 years. When the sampling began at 50 or 60 years, the graph was mainly linear. The decline in the last part of the age span (60–85 years) was almost identical regardless of the age at which sampling started, ranging between $z=-2.19$ and $z=-2.25$. The discrepancy between the quadratic and the smoothing spline was particularly pronounced in middle age, where the quadratic function would suggest a peak, while the local smoothing function would suggest a linear decline.

Comparison of AIC and BIC between models

For the wide age-ranges, the smoothing spline outperformed the quadratic models, with AIC Δ_I for the quadratic models of 16 and 9 for sampling from 8 and 20 years, respectively, and BIC Δ_I of 9 and 4. When sampling began at 30 years or higher, the models were not distinguishable with AIC $\Delta_I<2$. BIC tended to indicate a less good fit for the smoothing spline when sampling started at 30 or 40 years ($\Delta_I=3$ and 4, respectively).

Replication sample

To test the stability of the results, the same analyses were conducted in an independent sample of 309 participants. The lowest age in this sample was 18 years and the highest 94, so the age-ranges

Table 2
Top/break points of the curves, estimated age-decline and AIC.

Age span	Quadratic model				Local smoothing			
	Top point[a] (year)	Change 60–85 years (*z*-score)	AIC Δ*i*	BIC Δ*i*	Break point[c] (year)	Change 60–85[3] 85[b] years (*z*-score)	AIC Δ*i*	BIC Δ*i*
08–85	20	−1.64	16	9	49	−2.19	0	0
20–85	29	−1.87	9	4	49	−2.22	0	0
30–85	41	−2.26	0	0	49	−2.22	1	3
40–85	48	−2.50	0	0	47	−2.25	2	4
50–85	52	−2.59	0	0	Linear	−2.28	1	2
60–85	Linear	−2.62	2	1	Linear	−2.21	0	0

AIC: Akaike's Information Criterion.
BIC: Bayesian Information Criterion.
Δ*i* The difference between AIC/BIC for the model (*i*) and the model with the lowest AIC/BIC.
[a] The age at which the highest *y*-value was obtained.
[b] Change was calculated as the difference between estimated hippocampal volume (*z*-scores) at age 60 and age 85, based on the curve with the smoothing level that minimized AIC.
[c] The age where shows the age where the expression $-\frac{d^2f(age)}{d\,age^2}$ was largest, i.e. the point where the slope of the local smoothing curve changed the most (the second derivative), based on the curve with the smoothing level that minimized AIC.

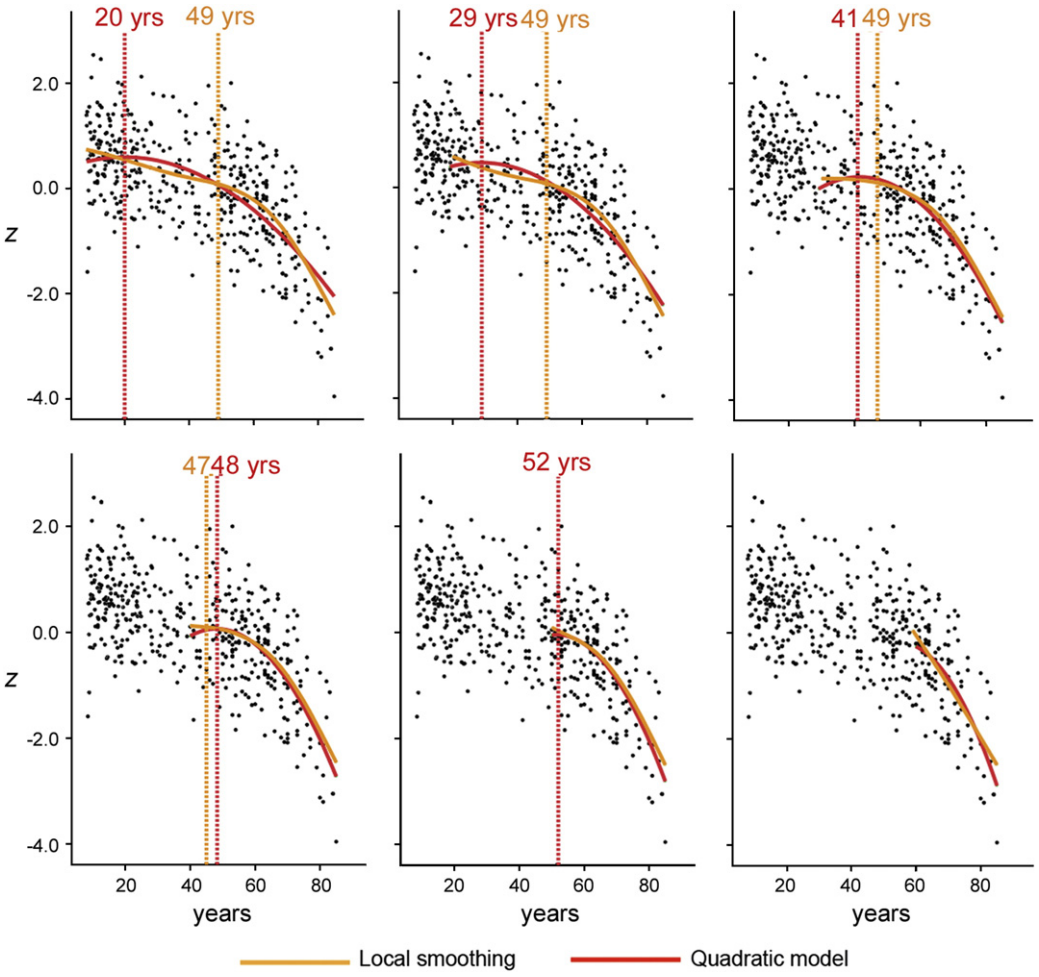

Fig. 3. Graphs of quadratic and local smoothing models at different starting ages. The red line represents the quadratic model and the yellow line the local smoothing model (the smoothing spline) for hippocampal volumes (standardized residuals after the effects of intracranial volume and sex were regressed out). The vertical red dotted line shows the peak of the quadratic function, while the dotted yellow line shows the age where the expression $-\frac{d^2 f(age)}{d\,age^2}$ was largest, i.e. the point where the slope of the local smoothing curve changed the most (the second derivative). Note that the quadratic model assumes the second derivative to be constant across the life-span, and hence the point of maximal acceleration cannot be determined. The quadratic model indicates a non-monotonic relationship between age and hippocampal volume. As can be seen, the estimated peak shifted towards a higher age when the lower limit of the age-span samples was increased, and the estimated decline in the latest part of the life-span (i.e. beyond 60 years) increased. The local smoothing model was more robust to the influence of sampling range.

analyzed were 18–94, 30–94, 40–94, 50–94 and 60–94 years. The regression equations and *p*-, *F*- and R^2-values are presented in Supplementary Table 1. Results from peak and slope analyses, as well as AIC and BIC, are presented in Table 3. The relationship between age and hippocampal volume was more linear than in the main sample, yielding somewhat smaller effects of varying sampling range. Still, the main results were replicated. For the quadratic model, whenever sampling was skewed towards a higher starting point, the age at

Table 3
Replication sample.

Age span	Quadratic model				Local smoothing			
	Top point[a] (year)	Change 60–85 years (*z*-score)	AIC Δ*i*	BIC Δ*i*	Break point[c] (year)	Change 60–85[3] 85[b] years (*z*-score)	AIC Δ*i*	BIC Δ*i*
18–95	33	−1.43	3	0	45	−1.36	0	0
30–95	43	−1.58	0	0	45	−1.37	2	3
40–95	41	−1.57	0	0	45	−1.38	1	2
50–95	Linear	−1.59	0	1	Linear	−1.39	1	0
60–95	Linear	−1.59	0	3	Linear	−1.37	0	0

AIC: Akaike's Information Criterion.
BIC: Bayesian Information Criterion.
Δ*i* The difference between AIC/BIC for the model (*i*) and the model with the lowest AIC/BIC.
[a] The age at which the highest y-value was obtained.
[b] Change was calculated as the difference between estimated hippocampal volume (*z*-scores) at age 60 and age 85, based on the curve with the smoothing level that minimized AIC.
[c] The age where shows the age where the expression $-\frac{d^2 f(age)}{d\,age^2}$ was largest, i.e. the point where the slope of the local smoothing curve changed the most (the second derivative), based on the curve with the smoothing level that minimized AIC.

which volumes peaked was moved upwards and the estimated decline in the last part of the age-span became larger. Peak estimated volume was 33 years when the model was tested with the full sample (18–94 years), and increased to 43 when the sample was restricted to the participants from 20 years. When sampling began at 40, the estimated peak was not further moved upwards, but was found to be at 41 years. Further, as seen for the main sample, the estimated age decline from 60 to 85 years increased when sampling included participants over 30 years only (from $z=-1.43$ to $z=-1.57$) compared with sampling that included all participants. Due to the more linear relationship between age and hippocampal volume, decline did not increase notably when sampling were restricted to higher ages than 30 ($z=-1.57$ when starting at 40 years compared to $z=-1.59$ when starting at 50 or 60 years).

Similar to the result from the analyses conducted with the main sample, the smoothing spline models were not affected by manipulations of the age-ranges. The point where the rate of change of the slope was highest was 45 years regardless of whether sampling started at 18, 30 or 40 years. When the sampling began at 50 or 60 years, the graph was mainly linear. The decline in the last part of the age span (60–85 years) was almost identical regardless of the age at which sampling started, ranging between $z=-1.36$ and $z=-1.39$. AIC differed less between the quadratic and the smoothing spline models in the replication sample than in the main sample, with $\Delta_I>2$ only when sampling started at 18 years (smoothing spline performed best). For all other age-ranges, $\Delta_I<2$. For BIC, $\Delta_I=3$ for the smoothing spline when sampling began at 30, otherwise $\Delta_I<2$.

Discussion

The present results demonstrated that when using a conventional quadratic model, the age-trajectories of hippocampal volume were systematically affected by the age at which sampling started. The top point where volume growth is ending and decline is beginning was systematically estimated to be later when the sampling started at a higher age. Further, estimated decline in the latest part of the life-span was steeper with a higher sampling start age. The nonparametric local smoothing approach (the smoothing spline) was more robust to the effects of different start ages.

It is tempting to use descriptions of the quadratic fits to infer timing of peaks of growth. As demonstrated, this peak will vary depending on the age-range of the study. If the sampling range started at 8 years of age, the break point was estimated to be 20 years, compared to 29 years when sampling from 20 years, and 41 years when sampling started at 30. This trend was confirmed in the replication sample. Thus, great caution should be exercised in inferring break points of age-trajectories based on quadratic models. Also, while the quadratic model indicated non-monotonic trajectories when sampling started at age 50 or earlier in the main sample and at age 40 or earlier in the replication sample, this pattern was not found for the local smoothing spline in either samples. This puts forth the question of whether non-monotonic relationships often reported in aging research (Allen et al., 2005; Jernigan and Gamst, 2005; Kennedy et al., 2008; Terribilli et al., 2009; Walhovd et al., 2005) to some degree are exaggerated by use of quadratic model fits. This question warrants further investigation.

The age-range sampled also systematically affected the estimated age-decline after 60 years in the quadratic models. Since the fit is global, each data-point will have an effect on the estimated fit for all other points. Thus, moving the starting point of sampling from 8 to 20 years affected the decline from 60 to 85 years from −1.64 to −1.87 *z*-scores in the main sample. The same tendency was observed in the replication sample, although the effects were far less dramatic due to a generally less non-linear relationship between age and hippocampal volume For the smoothing spline, different starting points did not change the estimation of the point where the change in the slope of the graph was largest in either of the samples. Further, the estimated age-decline from 60 to 85 years did not vary notably, ranging from −2.19 to −2.25 *z*-scores in the main sample and from $z=-1.36$ to $z=-1.39$ in the replication sample.

Analyses of goodness of fit showed that the smoothing spline outperformed quadratic models when the age-range was large. When sampling started at 8 years or 20 years in the main sample, and at 18 years in the replication sample, AIC was lower for the smoothing spline than the quadratic models ($\Delta_I>3$). When only the participants from 30 years and older were included, the models were indistinguishable with regard to goodness of fit ($\Delta_I<2$). These results indicate that the smoothing spline yields better goodness of fit than the quadratic model when the age-range in questions is large, but that the models perform equally well when the age-range is more restricted. However, the main problem with the quadratic model is not the goodness of fit per se, but that the results are difficult to interpret since they are substantially affected by variations in sampling range.

The implication of the present findings is not that authors using quadratic fits on their data necessarily have been erroneous. It is a valid approach to use the significance of the quadratic term to reject the null-hypothesis that a relationship is linear, and many researchers have not interpreted the shape of the quadratic function literally as a description of the true developmental or aging trajectory. For instance, Kennedy and Raz (2009) recently used quadratic terms to reject linear models, without making assumptions that these capture the true shape of the trajectories. Walhovd et al. (in press) used the same approach, and stated: "If a nonlinear (quadratic) component significantly increased the amount of explained variance (…) that does not mean that the exact quadratic fit shown depicts the true age function, and these fits should not be used to interpret the exact timing of peaks and dips in the age functions". Other researchers used quadratic fits to reject the linear model, and then went on to use more complex models to describe the trajectories, i.e. exponential models (Luft et al., 1999; Tamnes et al., 2010) or locally weighted least square algorithms (LOESS) (Westlye et al., In press). We would recommend that testing quadratic terms should be considered as a middle step in the model building process, but not necessarily as an end. If the quadratic term is significant, other models should be considered.

Still, when a quadratic function is published, it is very likely that many readers will interpret this as the true shape of the age-curve. In some cases the quadratic model will yield the best description of the data, but we argue that it is important to be aware of the factors related to age-range and sample composition when interpreting quadratic or other higher order polynomial models. It is easy to find examples in the brain-aging literature of statements based on quadratic fits that can easily be misinterpreted, even though the authors themselves did not interpret the quadratic fit to yield an accurate description of the true age-trajectory. A quadratic fit may adequately have described the age-trajectory in these cases, and some of the authors explicitly discussed other possible statistical models, but it is easy to see how a reader of these papers can get the impression that quadratic age-trajectories are common and accurate descriptions of structural brain aging.

This paper has two related messages: first, to say that a relationship is non-linear does not mean that it is quadratic. Non-linear trends can be constructed which are non-quadratic, and the choice when testing for non-linearity should not only be the addition of quadratic terms to the model. In the present paper, cross-sectional data were used to show this point, but it should be noted that longitudinal studies are not immune from similar problems. Studies with different starting points and different sampling follow-ups may yield different shapes of the age-curve. Age-related trajectories estimated from cross-sectional data are not necessarily predictive of individual age-trajectories, and a correct model would have to account for both intraindividual change over time and interindividual differences. Second, caution should be exerted in interpreting the

quadratic model as an actual description of age-trajectories of neuroanatomical structures. The nonparametric smoothing spline was less vulnerable to variations in sampling range, and yielded a more realistic description of age-trajectories. Irrespective of the shape of the underlying trend, with enough data, the nonparametric estimate will be close to the true curve. This flexibility is the result of computing regression estimates locally. The nonparametric estimate will be consistent for the underlying regression curve, whatever it is, as long as the bandwidth is reasonable. In contrast, a parametric model will only be correct when it is correctly specified. Thus one might argue that there is less practical risks involved when nonparametric techniques are used, because they will always be correct eventually, while parametric techniques will fail under certain very common conditions.

A cost of nonparametric techniques is that many degrees of freedom are lost in the process of controlling for curves of an a priori very unlikely shape. Thus, the rate of convergence of the estimated regression function is much slower than it would be for the right parametric model, leading to a loss in power. One consequence is that in finite samples, a parametric model which is reasonably accurate may outperform a nonparametric model. Further, nonparametric regression modeling is also more complicated than the parametric modeling. Adjustments for group effects can involve difficult considerations, and the smoothing parameters must be chosen. Too much smoothing could result in loss of important information (approaching a linear least squares model), while too little smoothing may result in large variance and poor generalizability of the model. In the present paper, we used an algorithm that selects the smoothing level that minimized AIC, thus alleviating the need for arbitrary choices of smoothing level. However, this approach did not work well when sampling started at 60 years, where no smoothing at all yielded the lowest AIC. Here, a greater smoothing level had to be used to yield a more realistic model of the relationship between age and hippocampal atrophy. For this age-range, a smoothing level chosen by minimizing BIC seemed more appropriate. Further, statistical inference from nonparametric models is also complicated by the absence of parameter estimates, which means that no prediction equation is defined and the relationship between x and y must be graphed (Andersen, 2009). Thus, even though the nonparametric smoothing spline is a useful tool to determine the shape of the relationship between a predictor and the response, it has also disadvantages compared to the simpler quadratic regression model. Choices related to nonparametric smoothing are discussed by Haerdle (1990). Mueller (1985) discuss the estimation of peaks, inflection points, and other functionals of the regression curve.

Disclosure statement

Anders M. Dale is a founder and holds equity in CorTechs Labs, Inc, and also serves on the Scientific Advisory Board. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. All other authors state that there are no actual or potential conflicts of interest. Appropriate approval and procedures were used concerning human subjects participating in the study.

Acknowledgments

Funding: The Norwegian Research Council (177404/W50 to K.B.W., 175066/D15 to A.M.F.), University of Oslo (to K.B.W. and A.M.F.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2010.01.061.

References

Allen, J.S., Bruss, J., Brown, C.K., Damasio, H., 2005. Normal neuroanatomical variation due to age: the major lobes and a parcellation of the temporal region. Neurobiol. Aging 26, 1245–1260 discussion 1279-1282.
Andersen, R., 2009. Nonparametric methods for modeling nonlinearity in regression analysis. Annu. Rev. Sociology 35, 67–85.
Berg, L., 1984. Clinical Dementia Rating. Br. J. Psychiatry 145, 339.
Berg, L., 1988. Clinical Dementia Rating (CDR). Psychopharmacol. Bull. 24, 637–639.
Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. NeuroImage 23, 724–738.
Courchesne, E., Chisum, H.J., Townsend, J., Cowles, A., Covington, J., Egaas, B., Harwood, M., Hinds, S., Press, G.A., 2000. Normal brain development and aging: quantitative analysis at in vivo MR imaging in healthy volunteers. Radiology 216, 672–682.
Du, A.T., Schuff, N., Chao, L.L., Kornak, J., Jagust, W.J., Kramer, J.H., Reed, B.R., Miller, B.L., Norman, D., Chui, H.C., Weiner, M.W., 2006. Age effects on atrophy rates of entorhinal cortex and hippocampus. Neurobiol. Aging 27, 733–740.
Fan, J., C., Z., J., Z., 2001. Generalized Likelihood Ratio Statistics and Wilks phenomenon. Ann. Stat. 29, 153–193.
Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.
Fischl, B., Salat, D.H., van der Kouwe, A.J., Makris, N., Segonne, F., Quinn, B.T., Dale, A.M., 2004. Sequence-independent segmentation of magnetic resonance images. NeuroImage 23 (Suppl. 1), S69–S84.
Fjell, A.M., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., Dale, A.M., Walhovd, K.B., 2009. High consistency of regional cortical thinning in aging across multiple samples. Cereb. Cortex 19, 2001–20012.
Fjell, A.M., Westlye, L.T., Greve, D.N., Fischl, B., Benner, T., van der Kouwe, A.J., Salat, D., Bjornerud, A., Due-Tonnessen, P., Walhovd, K.B., 2008. The relationship between diffusion tensor imaging and volumetry as measures of white matter properties. NeuroImage 42, 1654–1668.
Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage 14, 21–36.
Greenberg, D.L., Messer, D.F., Payne, M.E., Macfall, J.R., Provenzale, J.M., Steffens, D.C., Krishnan, R.R., 2008. Aging, gender, and the elderly adult brain: an examination of analytical strategies. Neurobiol. Aging 29, 290–302.
Haerdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.
Han, X., Fischl, B., 2007. Atlas renormalization for improved brain MR image segmentation across scanner platforms. IEEE Trans. Med. Imaging 26, 479–486.
Jernigan, T.L., Gamst, A.C., 2005. Changes in volume with age—consistency and interpretation of observed effects. Neurobiol. Aging 26, 1271–1274 discussion 1275-1278.
Kennedy, K.M., Raz, N., 2009. Aging white matter and cognition: differential effects of regional variations in diffusion properties on memory, executive functions, and speed. Neuropsychologia 47, 916–927.
Kennedy, K.M., Erickson, K.I., Rodrigue, K.M., Voss, M.W., Colcombe, S.J., Kramer, A.F., Acker, J.D., Raz, N., 2008. Age-related differences in regional brain volumes: a comparison of optimized voxel-based morphometry to manual volumetry. Neurobiol. Aging 30, 1657–1676.
Luft, A.R., Skalej, M., Schulz, J.B., Welte, D., Kolb, R., Burk, K., Klockgether, T., Voight, K., 1999. Patterns of age-related shrinkage in cerebellum and brainstem observed in vivo using three-dimensional MRI volumetry. Cereb. Cortex 9, 712–721.
Lupien, S.J., Evans, A., Lord, C., Miles, J., Pruessner, M., Pike, B., Pruessner, J.C., 2007. Hippocampal volume is as variable in young as in older adults: implications for the notion of hippocampal atrophy in humans. NeuroImage 34, 479–485.
Makris, N., Meyer, J.W., Bates, J.F., Yeterian, E.H., Kennedy, D.N., Caviness, V.S., 1999. MRI-Based topographic parcellation of human cerebral white matter and nuclei II. Rationale and applications with systematics of cerebral connectivity. NeuroImage 9, 18–45.
Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. 19, 1498–1507.
Mueller, H.G., 1985. Kernel estimators of zeros and of location and size of extrema of regression functions. Scand. J. Statist. 12, 221–232.
Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K.M., Williamson, A., Acker, J.D., 2004. Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. Neurobiol. Aging 25, 377–396.
Sowell, E.R., Peterson, B.S., Kan, E., Woods, R.P., Yoshii, J., Bansal, R., Xu, D., Zhu, H., Thompson, P.M., Toga, A.W., 2007. Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. Cereb. Cortex 17, 1550–1560.
Sowell, E.R., Peterson, B.S., Thompson, P.M., Welcome, S.E., Henkenius, A.L., Toga, A.W., 2003. Mapping cortical change across the human life span. Nat. Neurosci. 6, 309–315.
Sullivan, E.V., Marsh, L., Mathalon, D.H., Lim, K.O., Pfefferbaum, A., 1995. Age-related decline in MRI volumes of temporal lobe gray matter but not hippocampus. Neurobiol. Aging 16, 591–606.
Tamnes, C.K., Østby, Y., Fjell, A.M., Westlye, L.T., Due-Tonnessen, P., Walhovd, K.B., 2010. Brain maturation in adolescence and young adulthood: Regional age-

related changes in cortical thickness and white matter volume and microstructure. Cerebral Cortex 20, 534–548.
Terribilli, D., Schaufelberger, M.S., Duran, F.L., Zanetti, M.V., Curiati, P.K., Menezes, P.R., Scazufca, M., Amaro Jr., E., Leite, C.C., Busatto, G.F., 2009. Age-related gray matter volume changes in the brain during non-elderly adulthood. Neurobiol. Aging (Mar 10, Electronic publication ahead of print).
Van Petten, C., 2004. Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. Neuropsychologia 42, 1394–1413.
Walhovd, K.B., Fjell, A.M., Reinvang, I., Lundervold, A., Dale, A.M., Eilertsen, D.E., Quinn, B.T., Salat, D., Makris, N., Fischl, B., 2005. Effects of age on volumes of cortex, white matter and subcortical structures. Neurobiol. Aging 26, 1261–1270 discussion 1275-1268.
Walhovd, K.B., Moe, V., Slinning, K., Due-Tonnessen, P., Bjornerud, A., Dale, A.M., van der Kouwe, A., Quinn, B.T., Kosofsky, B., Greve, D., Fischl, B., 2007. Volumetric cerebral characteristics of children exposed to opiates and other substances in utero. NeuroImage 36, 1331–1344.
Walhovd, K.B., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D., Greve, D., Fischl, B., Dale, A., in press. Consistent neuroanatomical age-related volume differences across multiple samples. Neurobiol. Aging (2009 Jun 29, Electronic publication ahead of print).
Westlye, L.T., Walhovd, K.B., Bjornerud, A., Due-Tonnessen, P., Fjell, A.M., 2009. Error-related negativity is mediated by fractional anisotropy in the posterior cingulate gyrus—a study combining diffusion tensor imaging and electrophysiology in healthy adults. Cereb. .Cortex 19, 293–304.
Westlye, L.T., Walhovd, K.B., Dale, A.M., Bjørnerud, A., Due-Tonnessen, P., Engvig, A., Grydeland, H., Tamnes, C.K., Østby, Y., Fjell, A.M., In press. Life-span changes if the human white matter: diffusion tensor imaging (DTI) and volumetry. Cerebral Cortex (2009 Dec 23, Electronic publication ahead of print).