Published in final edited form as:

Behav Genet. 2009 September ; 39(5): 472–495. doi:10.1007/s10519-009-9272-1.

A Biometric Latent Curve Analysis of Memory Decline in Older Men of the NAS-NRC Twin Registry

John J. McArdle and
University of Southern California

Brenda L. Plassman
Duke University Medical Center

Abstract

Previous research has shown cognitive abilities to have different biometric patterns of age-changes. Here we examined the variation in episodic memory (Words Recalled) for over 6,000 twin pairs who were initially aged 59–75, and were subsequently re-assessed up to three more times over 12 years. In cross-sectional analyses, variation in Education was explained by strong additive genetic influences (~43%) together with shared family influences (~35%) that were independent of age. The longitudinal phenotypic analysis of the Word Recall task showed systematic linear declines over age, but with positive influences of Education and Retesting. The longitudinal biometric estimation yielded: (a) A separation of non-shared environmental influences and transient measurement error (~50%); (b) Strong additive genetic components of this latent curve (~70% at age 60) with *increases over age* that reach about 90% by age 90. (c) The minor influences of shared family environment (~17% at age 60) were effectively eliminated by age 75. (d) Non-shared environmental effects play an important role over most of the life-span (peak of 42% at age 70) but their relative role diminishes after age 75.

Keywords

Longitudinal structural equation modeling; multi-level mixed-effect biometric models; multivariate twin and family data analysis; cognitive aging; memory loss

INTRODUCTION

Questions about cognitive development at any point in the life-span can be examined using cross-sectional data on aging individuals. The inclusion of family data, especially twin data, has allowed separation of the relative impact of genes and environments on individual differences (e.g., McClearn et al, 1997). But it is clear that longitudinal data is required to measure age-related changes and provide the initial empirical basis for inferences on the sources of these changes (e.g., Baltes & Nesselroade, 1979). It follows that longitudinal information about individual trajectories can be further enhanced by collecting data on related members of a family, and longitudinal data collected on pairs of twins can be especially informative (e.g., Blum & Jarvik, 1974; McArdle et al, 1998; Reynolds et al, 2005).

The application of developmental-genetic concepts to issues in adult cognition often starts with a representation of *memory loss* (e.g., Brown, 1974; Hultch et al, 1998; Zachs & Hasher in

Bialystok & Craik, 2006). It is clear that memory loss is one of the most important descriptors of human cognitive aging: "Some slight decreases in long-term memory efficiency are widely reported in all studies of old people, and probably reflect exactly what is happening in the daily lives of their subjects. ... Also, in contrast to the dementias, little is known about the brain processes that cause mild types of memory impairment. The ignorance of the biological basis of mild memory impairment is the major obstacle preventing proper drug development in this area." (Whalley, 2001, pp. 73-74). Many researchers suggest that adult memory losses at some earlier age act in some way as a precursor or at least as a leading indicator of more severe forms of brain damage (e.g., Petersen et al. 1999; Katzman *et al.*, 1989)

The biometric basis of memory loss has been studied as a key part of other longitudinal twin and family studies. The initial studies of Kallman and Jarvik with the New York Twin samples ($N=268$; see Blum & Jarvik, 1974) found high heritability for memory span (46%) and a strong association between performance on a memory task and education, but did not examine rates of change over age. The recent cross-sectional study of SATSA twins aged 80 or older by McClearn et al (1997; $N=240$ pairs) again found high heritability of memory (52%). In contrast to the previous studies, the cross-sectional study of Brandt et al (1993) assessed more than 4,000 aging twin pairs from the Duke Twins Study (see below) and again found education effects but lower heritability for delayed recall (20%). Longitudinal twin studies of many different cognitive functions have been studied (e.g., Pedersen et al, 1992). The most recent longitudinal studies of Swedish Twins (SATSA; see Finkel et al, 2003, 2005; Pedersen et al, 2003; Reynolds, 2002, 2005) all found substantial additive genetic variance for baseline memory performance, but little genetic contribution to memory decline over time.

There are many methodological questions raised by developmental genetics. These issues were clarified in classic work of Vandenberg & Falkner (1965) who showed how the biometric basis of a polynomial growth curve model could be calculated. This broad class of math-stat models was later represented in the structural equation form of a *latent curve model* with biometric components by McArdle (1986). Subsequent research has improved upon these methods (e.g., Neale & McArdle, 2000; McArdle & Hamagami, 2003; Neale et al, 2006; McArdle, 2006) and applied these methods to substantive problems (e.g., McArdle, et al, 1998; Finkel et al, 2003; McGue & Christensen, 2002; Reynolds et al, 2002, 2005).

We used longitudinal twin analyses to study the variation in episodic memory (based on immediate and delayed Word Recall tasks) for over 6,000 pairs of male twins who were initially aged 59-75 at the first testing, and then were re-assessed up to three more times over a 12 year period. The large-scale collection of longitudinal data on twins is rare, and it permits an in-depth evaluation of the genetic and non-genetic influences on latent changes in memory loss in the elderly. To analyze these data, we used a new approach to longitudinal biometric analysis based on mixed-effects models from McArdle & Prescott (2005). We also followed the longitudinal approach presented by McArdle (2006) and used orthogonal variance components for the biometric decomposition of components of change in longitudinal twin analysis. Using this specific approach it is relatively straightforward to account for incomplete longitudinal or twin pair data, and we can examine attrition over time and within pairs using *MLE-MAR* incomplete data techniques for sampling bias correction (e.g., McArdle, Prescott, Horn & Hamagami, 1998; cf., Pedersen, Ripatti, Berg, Reynolds, Hofer, & Finkel, 2003). This approach also easily permits the examination of a wide variety of latent basis curves, such as the use of *age-at-testing* instead of *wave-of-testing*, and these are important choices in substantive applications of these techniques to cognitive aging (e.g., McArdle, et al, 1998; Finkel, Reynolds, McArdle, Gatz & Pedersen, 2003).

METHODS

Participants

Participants were members of the Duke Twins Study of Memory in Aging in the National Academy of Sciences-National Research Council (NAS-NRC) Twin Registry of WWII Veterans (Jablon et al, 1967; Page, 2002; Plassman et al, 2006) who completed the self-report telephone cognitive screening measure at least once during Waves 1- 4 of data collection. As part of the Duke Twins Study, members of the NAS-NRC Twin Registry, born between 1917 and 1927, have been studied over four longitudinal waves of telephone screening and in-person clinical assessment for dementia. The data collection was described in detail by Plassman et al (2006). For the majority of twin pairs, zygosity was determined by questionnaire, from military records (physical characteristics such as height, weight, eye and hair color), fingerprint records, and (for a sample) blood group testing. The items used in the questionnaire were essentially translations of those used in the Swedish and Danish twin registries. This method of establishing zygosity has been estimated by cross-validation with DNA to be 97 percent accurate. (Reed et al., 2005) For those individuals who were assessed in-person due to suspected dementia and their cotwins, we routinely obtained blood or buccal DNA samples to determine zygosity.

Trained lay interviewers administered a self-report (described below) or proxy cognitive screening measure to N=13,532 individuals up to four time over a period of 12 years (see Table (1)). Individuals targeted for participation at each wave were twin pairs in which neither twin was identified as demented at previous waves and in which both members were still thought to be alive. The exception to this protocol was that in Wave 4, we attempted to assess as many of the singletons (living member of pairs in which one twin was known to be deceased) as possible. The exclusion of singletons at other screening waves was due to practical rather than statistical concerns, and we discuss these selection issues as part of the longitudinal analyses to follow.

Cognitive Measurements

The original *Telephone Interview for Cognitive Status* (TICS) instrument (Brandt, 1988) and its modified form, the TICS-m (Welsh et al, 1993) were developed to screen for dementia and to provide a brief assessment of cognitive status that could be administered over the telephone. The TICS-m has a maximum of 50 points and is highly correlated with the Mini-Mental State Examination (Plassman, 1994). The TICS-m included items assessing the following cognitive domains: orientation, concentration, memory, naming, comprehension, calculation, and reasoning. The memory items included immediate and delayed recall of a list of 10 words (for details, see Welsh, et al, 1993).

There are many possible combinations of these cognitive variables that could be reasonable for further study. One principal components analyses of the Wave[1] data (Brandt et al 1993) as well as a common factor structural analyses of related TICS data (HRS; see McArdle, Kadlec & Fisher, 2007) yielded similar results. These studies showed that the two word recall scores, Immediate and Delayed, can be isolated from the other cognitive variables to form a separate factor of Episodic Memory. This factor is indicated here by a composite score (Word Recall score) formed as the average of the two scores (0-10 words per task) multiplied by 10 for easy computations and interpretation as a *percent correct* scale (0-100% of task).

Longitudinal Data Description

An overall description of the Duke Twins Study sample data used in these analyses is presented in Table(1). The total number of individual participants ($N=13,532$) is separated into specific sub-groups describing their patterns of measurement: (a) at all four waves ($n=3,948$, 29.2%),

(b) at three waves ($n=2,638$, 19.5%), (c) at two waves ($n=4,225$, 31.2%), and at only one wave ($n=2,721$, 20.1%). We also present information on the number of total measurements ($M=34,874$), and the parts contributed by each subgroup, and here the subgroups contribute more if they are measured on more occasions (i.e., the subgroup that completed all four waves of screening represent 45.3% of all measurements). The subgroups also differ on their mean Educational level, their average Age at the first wave of testing, and their average score on the Word Recall task (described above). Although these differences appear small, even small differences can lead to biases because the complete cases (first group) report higher education, lower age, and higher Word Recall. The problems this can create for data analysis are described later, but in general, we try to overcome these kinds of biases by using all the available measurements in all analyses.

In Table (2) the same longitudinal data are reorganized into MZ and DZ groups. These are presented here as descriptive sampling statistics because no restrictive biometric model has yet been applied. Of course, it is clear that the intra-class correlations are higher for the MZ than the DZ pairs.

Selected data on the NAS-NRC sample at Wave[1] are also presented in Figure (1). The data included here are based on $N=6,197$ pairs who have valid information on the TICS-m. The mean age at Wave[1] for all persons is $\text{Age}[1]=66.5$, with a range of 59-75, and the frequency distribution of Figure (1a) is skewed, with many more participants at the youngest ages (<65). Figure (1b) is a plot of the Word Recall Score showing a nearly symmetric distribution with a range from 0-100% and a mean of 29.3% ($sd=14.1$). This translates to approximately 3 words recalled per task. Figure (1c) is a side-by-side boxplot illustrating the small negative relationship between Word Recall Score and Age (correlation = -0.14), suggesting lower scores after age 65-70.

Figure (2) is a plot of individual trajectories of the Word Recall Score for a 5% random sample based on only one person from each family. From 1 to 4 scores for each person were obtained at different time points, and these individual data are connected by a line so that more of the entire pattern of change can be visualized. Figure (2a) is a plot of individual trajectories in Word Recall over 4 *Waves of Testing*. The limited number of scores yields gaps at each wave, but no discernable pattern of changes over waves can be seen. In contrast, Figure (2b) uses the same data, but here we plot the individual trajectories in Word Recall over the individual's *Age-at-Testing*. Although this picture is also fairly complicated, we now can see the more typical tendency for the scores to decline over age (e.g., McArdle et al, 2007).

Mixed-Effect Biometric Variance Component Models

Our analytic approach follows McArdle & Prescott (2005) who showed how simultaneous estimation of the biometric parameters can be programmed using current MEMA programs (e.g., SAS PROC MIXED). The benefits of writing the model using structural equation techniques include *maximum-likelihood estimation (MLE)* and standard indices of goodness-of-fit (i.e., χ^2 , df , ϵ_a). This analytical approach is based on the communalities among *path analysis models* (PAM) and *variance component models* (VCM), and mixed-effects models can yield identical information as well as biometric inferences. This approach can easily include measured covariates, observed variable interactions, and multiple relatives within each family, and the programming logic is relatively standard. For clarity here, we briefly review the basic elements of this approach.

We denote the observed score (Y_n) for any person ($n=1$ to N) but include different families ($f=1$ to F) and different persons ($p=1$ or 2) within the family. We then write a structural equation model for any pair of persons as

$$\begin{aligned}
 Y_{f,1} &= \mu + A_{f,1} + S_f + I_{f,1} = \mu + \sigma_a a_{f1} + \sigma_s s_f + \sigma_i i_{f1}, \text{ and} \\
 Y_{f,2} &= \mu + A_{f,2} + S_f + I_{f,2} = \mu + \sigma_a a_{f2} + \sigma_s s_f + \sigma_i i_{f2}, \text{ with} \\
 E\{a_1, a_2\} &= \rho_a.
 \end{aligned}
 \tag{1}$$

In this expression we represent the mean (μ) independent sources of the deviations that are additive genetic (A), non-genetic but shared by family members (S), and non-genetic factors independent across individuals (I). (Note: We recognize the terms A , C , and E are more typical labels for the same biometric scores). These deviations can also be written with unobserved scores scaled to unit variance ($E\{a, a\} = E\{s, s\} = E\{i, i\} = 1$) so the coefficients (σ_j) represent the standard deviation of each component, and the genetic correlation is assigned based on the genetic relationship of the pair (e.g., for MZ pairs, $\rho_a = 1$, whereas for DZ pairs, $\rho_a = 1/2$). In the simplest version of biometric theory we assume these components are all uncorrelated. More complex versions of this model include components to represent non-additivity (e.g., dominance deviations), examine correlations among these components (e.g., $|\rho\{A, I\}| > 0$), and consider interactions among components (e.g., A by I).

The standard model for a pair of relatives is often drawn as a latent variable path diagram that is familiar to behavioral genetics (BG) researchers. As shown in detail by McArdle & Prescott (2005), some useful features emerge when we restrict our expression of any BG model so “all paths are fixed values for each person.” We can rewrite Equation (1) for a pair of persons within the same family as

$$\begin{aligned}
 Y_{f,1} &= \mu + A_{f1} + S_f + I_{f1}, \text{ with } A_{f1} = w_{ac} AC_f + w_{au} AU_{f1}, \text{ and} \\
 Y_{f,2} &= \mu + A_{f2} + S_f + I_{f2}, \text{ with } A_{f2} = w_{ac} AC_f + w_{au} AU_{f2},
 \end{aligned}
 \tag{2}$$

where the additive genetic deviation is separated into two deviation scores: (a) AC_f is common for members of the same family, and (b) AU_{fp} is unique to the person. In this form the weights (W) are fixed at values which indicate the proportion of the additive genetic deviation shared between relatives. This general variance components model is drawn following Equation (2) as a nested or higher-order latent variable or reduced form path diagram (see Figure (1) in McArdle & Prescott, 2005).

The computational approach suggested by McArdle & Prescott (2005) was based on this reparameterization of the usual path model so the paths (coefficients) are all fixed weights defined by the application. These weights are typically scaled so the sum of squares is unity ($w_{ac}^2 + w_{au}^2 = 1$) so they do not impact the variance terms. Since the two individuals in any MZ pair are assumed to share the same genotypes and the same common family environment, we simplify the model for an MZ pair by fixing $w_{ac} = 1$ and $w_{au} = 0$ (in Eq. (2)). In contrast, the assumption of no assortative mating implies that members of DZ pairs share $1/2$ of their genotypes (due to segregation of alleles), and this implication can be represented as fixed weights of $w_{ac} = \text{sqrt}(1/2)$ and $w_{au} = \text{sqrt}(1/2)$. In order to estimate genetic dominance we add two new latent scores, D_{fp} for each person, decomposed into common DC_f and unique DU_{fp} , with both assigned equal variance (i.e., σ_d^2). The analysis of MZ-DZ twins would require corresponding weights of 1 and 0 in MZ pairs for DC and weights of $\text{sqrt}(1/2)$ for both unique DU_{fp} .

Latent Curves as Mixed Effects Models

We now extend the path analysis and variance components models to a longitudinal curve model for twin data following the techniques in McArdle (1986, 2006). In these analyses we are interested in the trajectory-over-time as well as the nested structure of the correlations

among relatives, so we present the model needed to estimate all of the parameters of the model – both longitudinal and biometric. One generic form of a *latent curve model* based on a trajectory over time for observed variables is written as

$$Y[t]_n = g0_n + \mathbf{B}[t] g1_n + u[t]_n, \quad (3)$$

where for any individual ($n=1$ to N) we write three unobserved latent variables: (a) $g0$ are termed the initial level scores; (b) $g1$ are termed the slope scores, and (c) $u[t]$ are unobserved and independent unique errors for measurements at each occasion. In some models there are assumed to be random errors, but in other models they also contain either systematic specific tests or systematic state variation as well (e.g., McArdle & Woodcock, 1997). The set of basis coefficients $\mathbf{B}[t]$ represents a function of the timing of the observations. In some cases $\mathbf{B}[t] = [0, 1, 2, 3]$ represents the retesting wave, or we can write $\mathbf{B}[t] = f\{\text{Age}[t]\}$ to represent the age at the time of the observation (see Figure (2b) in McArdle et al, 1998).

The typical application of this latent growth model further presumes that the initial level and slopes are assumed to be random variables with “fixed” means (μ_0, μ_1) but with unobserved deviations ($d0, d1$) that have “random” variances (σ_0^2, σ_1^2) and covariance (σ_{01}). The typical path diagram gives a standard representation of this kind of latent growth model. Model parameters representing “fixed” or “group” coefficients are drawn as one headed arrows while “random” or “individual” features are drawn as two-headed arrows. As usual, the unique error terms are assumed to be normally distributed with mean zero and constant variance (σ_u^2) and are presumably uncorrelated with all other components. It is also possible to rewrite the standard model adding a new latent score $c01$, which is interpreted as a common factor of level and slope, with uncorrelated unique scores ($u0, u1$). Now by simple substitution, we can rewrite the reduced form of the basic growth model as

$$\begin{aligned} Y[t]_n &= (c01_n + u0_n) + \mathbf{B}[t] (c01_n + u1_n) + u[t]_n \\ &= u0_n + \mathbf{B}[t] u1_n + c01_n + \mathbf{B}[t] c01_n + u[t]_n \\ &= u0_n + \mathbf{B}[t] u1_n + (1 + \mathbf{B}[t]) c01_n + u[t]_n, \end{aligned} \quad (4)$$

where there are two sets of basis coefficients ($\mathbf{B}[t]$ and $1 + \mathbf{B}[t]$) but all the model components ($c01, u0, u1, u[t]$) are presumed to be uncorrelated. As it turns out, the orthogonal reduced form of the latent curve model written as Equation (4) is especially useful for representing correlated latent variables as orthogonal variance components calculations (e.g., SAS MIXED).

Longitudinal Biometric Variance Component Models

We next represent the biometric decomposition of the latent components of level and slope following McArdle (1986). First, we assume the longitudinal growth representation of Equation (3) is appropriate for all individuals. Next, using pairs of relatives, we consider the initial level ($g0$) to be decomposable into three new latent variables that are added within each person ($A0_n, S0_f, E0_n$) and where the genetic correlation is assigned for the group (e.g., $\rho_a=1$ or $\rho=1/2$). In this way, the biometric parameters ($\sigma_{0a}^2, \sigma_{0s}^2, \sigma_{0e}^2$) indicate features of the latent initial level variance. We write the same model for the latent slopes ($g1$) across pairs and include a regression on the initial level latent variables, although we recognize this choice of the “slope as an outcome” is usually arbitrary. A path analysis diagram of this model is presented in Figure (3) and, as shown in many other research reports, this model can be fitted as a structural equation model with multiple groups.

To fit this same model using mixed-effects techniques we can use the orthogonal variance components approach established in Equation (4). We re-parameterize both common and

unique parts of these components to specify the model as a set of orthogonal variance components, and we rewrite the model in terms of fixed weights and in the reduced form of

$$\begin{aligned} Y[t]_{f,p} &= u0_{f,p} + \mathbf{B}[t] \cdot u1_{f,p} + (1 + \mathbf{B}[t]) \cdot c01_{f,p} + u[t]_{f,p}, \\ u0_{f,p} &= \mu_0 + Au0_{f,p} + Su0_{f,p} + Iu0_{f,p}, Au0_{f,p} = w_{ac}ACu0_{f,p} + w_{au}AUu0_{f,p}, \\ u1_{f,p} &= \mu_1 + Au1_{f,p} + Su1_{f,p} + Iu1_{f,p}, Au1_{f,p} = w_{ac}ACu1_{f,p} + w_{au}AUu1_{f,p}, \text{ and} \\ c01_{f,p} &= Ac01_{f,p} + Sc01_{f,p} + Ic01_{f,p}, Ac01_{f,p} = w_{ac}ACc01_{f,p} + w_{au}AUc01_{f,p}. \end{aligned} \quad (5)$$

In this model the additive genetic weights are defined based on the genetic resemblance of the pairs (e.g., $W_{mz} = [1, 0]$, $W_{dz} = [\text{sqrt}(1/2), \text{sqrt}(1/2)]$), so we estimate two means, one error variance, and nine orthogonal variance components (three which represent covariance information). The higher-order version of the orthogonal model is portrayed in McArdle (2006) and we can now fit the biometric latent curve models to these data using any of several programs (SAS, Mplus, Mx) and obtain the same basic results (all programs used here are available from the first author).

RESULTS

Analyses 1: Cross-Sectional Biometric Results for Educational Attainment

The first series of analyses were based only on the cross-sectional data at Wave[1] on self-reported Educational Attainment with the final results listed in Table (3). These models included biometric parameters that were estimated from the basic mixed-effects model (Eq. 2, in SAS PROC MIXED and NLMIXED) using data from all twin pairs from their first assessment ($N=6,488$ twin pairs with $D=11,796$, possibly from different waves).

Initially, a baseline model was fit with a mean and an independent error term only ($L^2=65288$ with 2 parameters; but not listed in Table 2). The second model added a shared family component, and this improved the fit ($\chi^2=3245$ on $df=1$). The third model included an additive genetic component, and this also improved the fit ($\chi^2=4755$ on $df=1$). The overall misfit of this biometric restricted model is very small (by comparison to L^2 in Table 2, $\chi^2=19$ on $df=4$). Of course, there are age differences between families at the initial testing, the addition of age as a linear covariate in this Education model improved the fit by a significant amount ($\chi^2=67$ on $df=1$), but age accounted for only a small percentage of the Education variation.

Parameters and z-values for the two final models are presented in Table (3). In column (3b) the mean parameters include an intercept at Age 70 ($\beta_0=12.85$ in years), and a small but significant linear decline in Educational Attainment with age ($\beta_1=-.54$ per decade). The variance parameters include the contribution of additive genetic effects ($\sigma_a^2=4.3$), of non-age related shared family genetic effects ($\sigma_s^2=3.48$), and of independent non-shared family genetic effects ($\sigma_u^2=2.27$). If we ignore the age related variance (5%), the remainder of the variance can be written as percentages of a total variance of $\rho_a^2=42.8\%$ (often referred to as h^2), $\rho_s^2=34.6\%$, and $\rho_e^2=22.6\%$. While Educational Attainment is largely heritable, there are large common family and individual components. The results of the final biometric structure of Educational Attainment are presented as a path diagram in Figure (4).

Analysis 2: Sequential Cross-sectional Biometric Results for Word Recall

The next series of analyses were based only on the cross-sectional data at each wave (Wave [1] to Wave[4]) for Total Word Recall, and numerical results are presented for each wave in Tables (4) and (5). The first column in Table (4) is a decomposition of Word[1] for $N=6,440$ twin pairs. Once again, a baseline model was fit with one mean and one independent variance only. The second model added the shared family component, and this improved the fit ($\chi^2=319$ on $df=1$). The third model included an additive genetic component, and this also improved the

fit ($\chi^2=3,299$ on $df=1$). Once again, the overall misfit of this biometric restricted model is very small ($\chi^2=29$ on $df=4$).

Models with covariates of Age and Education are included in Table (5). In the case of Word Recall at Wave[1], the addition of Age as a linear covariate in the model also improved the fit ($\chi^2=233$ on $df=1$), and then Educational level was added as well ($\chi^2=107$ on $df=1$). The results of the final biometric structure of Word Recall at Wave[1] is presented in Figure (5). The mean parameters listed here include an intercept at Age 70 ($\beta_0=25.4$), and a linear decline with age ($\beta_1=-6.6$ per decade). The variance parameters include the contribution of additive genetic effects ($\sigma_a^2=33.1$), of non-age related shared family genetic effects ($\sigma_s^2=6.5$), and of independent non-shared family genetic effects ($\sigma_e^2=132.4$).

If we ignore the age related variance (7%), the biometric variance of Word[1] can be decomposed as percentages as $\eta_a^2=22\%$ (h^2), $\eta_s^2=11\%$, and $\eta_e^2=67\%$. It is important to note that in these univariate analyses this estimate for the non-shared variance includes whatever might be attributable to measurement error, so this large portion of non-shared variance probably does not accurately represent the biometric concepts intended (see McArdle & Goldsmith, 1990).

To get an initial idea about the likely age related changes to expect in the subsequent longitudinal models, we carried out the previous cross-sectional analyses within each wave (i.e., presented in the other columns of Tables (4) and (5)). Although these estimates were calculated separately and without cross-time restrictions, the biometric results for Word[2], Word[3], and Word[4] exhibited the same basic structure as found in Word[1] above. One unusual result in Table (4) was that there seemed to be a complete elimination of the shared family variance when we examined the last wave with Word[4]. The fit of all models was improved a great deal when Age and Education are included. In Table (5) we see consistent and large negative effects of Age on Word Recall (-8% per decade) together with large positive effects of Education on Word Recall ($+1\%$ per year). When this family and individual variation was taken into account, the residual cross-sectional results were remarkably consistent across waves – there was no significant shared variation, the additive genetic variation was small ($\sim 20\%$), and most of the individual variation was due to independent environments or measurement error ($\sim 80\%$).

As another cross-sectional view of the age-related longitudinal changes, the same kind of cross-sectional analysis was repeated for many different groups with “similar ages.” We created subgroups by centering on a specific age, and then considered individuals and pairs who were within two years of that age. Using sampling weights based on a normal smoother ($W=[.04, .19, .50, .19, .04]$) we could assure that all statistical information was based on a reasonable sample size. This exploratory analysis is conceptually simple, and we simply invoked the SAS Macro language and placed a DO-Loop around the SAS PROC MIXED code used above without the age covariate. The overall result of this sequential set of cross-sectional biometric analysis showed substantial variations in the additive variance within each age, not very much shared family variance within each age, and large non-shared variance (plus measurement error) within each age. Unfortunately, these sequential cross-sectional genetic estimates did not show distinct pattern over ages.

Analysis 3: Longitudinal Phenotypic Results for Word Recall

A linear latent curve twin model was first fitted using standard mixed-effect techniques in SAS PROC MIXED software and the results are presented in Table (6). These analyses included only one member of each twin pair, but included the participants no matter how many waves they were assessed. The algorithm used the maximum likelihood estimation, so these results are considered appropriate under “missing at random” (MAR) assumptions about the

incomplete cases – that is, the available data for any individual provides information about why the other data are missing (e.g., McArdle & Hamagami, 1992; Hedeker & Gibbons, 1997). These standard MAR assumptions are non-trivial here, because many of the incomplete cases were selectively due to illness or mortality, and we return to this issue again later. To examine the potential estimation biases due to selection, constraints were first fitted using the means only, then as a more complete model for both the means and covariances.

Three models for the phenotypic structure of Word Recall over Age were fitted with: (a) linear “age at testing” as the only predictor, (b) two additional covariates (retest and education), and (c) with additional interactions. The age-based model (6a) was used as the starting point, and shows significant linear age decline ($\beta_1 = -3.2$ per decade) with significant age-slope variation ($\sigma_1^2 = 4.9$). The second model (6b) included Education as a covariate, and fit much better than the first ($\chi^2 = 5616$ on $df = 3$), but the fit was not improved much by adding an impact of covariates on the age slopes ($\chi^2 = 41$ on $df = 3$). The additional model (6c) included the impact of Retest – taking the test at least once before – and this improved the fit substantially, but no slope impact or interactions were required. In the final model (6c) the key mean parameters are listed as: (1) the group intercept at age 70 ($\mu_0 = 26.1$), (2) the group linear age slope ($\mu_1 = -4.7$ per decade), with (3) significant fixed effects for Retesting (+2.3 for having taken this test at least once), and Education (+1.5 per year). The results for the variance parameters are separately listed as (4) the unique intercept at age 70 ($\sigma_{u0}^2 = 76.1$), (5) the unique linear age slope ($\sigma_{u1}^2 = 5.7$), (6) the correlation of intercept and slope ($\rho_{01} = .77$), and (7) a unique error variance ($\sigma_u^2 = 127$). When the common and unique parts are added together, we can make a prediction about the variation at any time point (but it changes over time). Here we list the expected contributions at age 80 and estimate 32% is from the overall time-invariant difference between people (intercept), about 8% is due to the systematic time-changing (slope) variance, while about 48% unique, combined specific variance and measurement error. In addition about 9% is due to educational differences, and 3% due to retest effects. Of course, these systematic change components look relatively small compared to the large error variance, but the ratio of the systematic to unique components changes over time and this longitudinal variation increases over age – the intra-class within time coefficient changes from $\eta[80]^2 = 0.3$ to $\eta[90]^2 = 0.6$ (see McArdle & Woodcock, 1997).

A variety of other nonlinear models were examined, including quadratic, two part splines, and exponential decline curves (see McArdle et al, 2002). These nonlinear models for the mean changes over age all fit slightly better than the simple linear models, and each indicated that the means were lower after age 75. However, the addition of variance components for these models did not indicate substantial improvement, so we simply retained the simple linear models for further evaluations. We also compared the parameter estimates and choice of models using only the cross-sectional sample, but these latent curve results led to essentially the same conclusions.

Analysis 4: Longitudinal and Biometric Models for Word Recall

The longitudinal biometric results presented in Table (7) include all participants in Table (1) (i.e., no matter how many measured waves). The first model includes the linear age curve with a full biometric decomposition (Figure 3). The results for the fixed parameters are, not surprisingly, similar to that found in phenotypic model (6a) -- the group intercept at age 70 ($\beta_0 = 25.6$), with a negative group linear age slope ($\beta_1 = -5.6$ per decade), and with positive Education ($\beta_e = +1.5$ per year) and Retest ($\beta_r = +2.9$ if not first test). These fixed effects have a similar pattern as found in the phenotypic model (6b) with positive impacts on the level of Word Recall.

The results for the variance parameters are separated into (1) unique error variance ($\sigma_u^2 = 124$), (2) the unique intercept at age 70 ($\sigma_{u0a}^2 = 28.1$, $\sigma_{u0s}^2 = 26.2$, $\sigma_{u0i}^2 = 6.6$), (3) the unique linear age

slope ($\sigma_{u1a}^2=9.5$, $\sigma_{u1s}^2=-16.9$, $\sigma_{u1i}^2=0.8$), and (4) the common variance of intercept and slope ($\sigma_{ca}^2=11.5$, $\sigma_{cs}^2=6.7$, $\sigma_{ci}^2=-1.1$). Negative parameter estimates were allowed, as these are not variances but components of variances. That is, these variance components can be combined to estimate the contributions to any latent score in the overall model (see McArdle, 2006), and the final variances they imply are presented in the path diagram of Figure (6). In order to examine a specific pattern of time-based effects, the second model (7b) restricted the linear slope over age to be entirely additive genetic components, and this model did not yield any negative variance estimates and fit the longitudinal twin data without substantial loss of fit ($\chi^2=14$ on $df=4$).

In one additional analysis we estimated the same longitudinal biometric model using only cases with complete data (see Table (1)). This reduced the available data (to 45%), but still affords a relatively large sample size ($n=3,948$ at $T=4$ waves). Although the means were slightly higher, the use of the complete case data only led to small changes in the parameter estimates and subsequent interpretations.

Analysis 5: Considering Changes Over Age in Word Recall

The final calculation of the expected mean and variance changes in Word Recall over age is presented in Table (8) and Figure (7). The group mean changes can be done from the previous parameters, and the result is a decline over age from $\mu[60]=29.3$ to $\mu[90]=15.4$. This assumed an initial test for persons with 12 years of Education, but adjustments in these averages can be made for other patterns of covariates. The appropriate calculation of proportions of variance can be a more complex problem when dealing with different time points (see McArdle, 1986, 2006; Eaves et al, 1986; Hewitt et al, 1988; Loehlin et al, 1989). One issue that needs to be considered is the changing baseline of expected variance at each age, and another issue is the addition of the independent and common components of the slope combined to form the change variance. The way we have dealt with this problem in prior research is to compare the changes in the expected deviations at specific ages (see McArdle, 1986; McArdle et al, 1998; McArdle & Hamagami, 2003; McArdle, 2006).

The first set of columns (8a) shows the estimated phenotypic changes over the three decades between ages 60 and 90 – these include a decreasing mean from 29 to 15, together with an increasing deviation of the no-error scores from 7 to 14. Under the assumption that the error variance is constant over time (an assumption of all the models used here), the expected reliability of the Word Recall measurement increases over time from .30 to .60. The second set of columns shows the biometric decomposition to the deviations over age. These are calculated from the variance component estimated from the final model (7a). The results show the increase in the additive genetic variance (more than doubling from 6 to 13), a smaller but steady contribution of the independent environments (3 to 6), but a small and decreasing impact of the shared family environments over age. These parameters are plotted over age in Figure (7). The final set of columns (8c) recasts these raw deviations as percentages of the total variation at each age, and this basically mimics the pattern of the raw deviations. However, now it is clear that by age 90 the increasing additive genetic variance over age overwhelms all other sources of variation considered here. That is, the estimate of the additive genetic variance of the true growth-decline in Word Recall is over 90% by age 90.

Analysis 6: Alternative Longitudinal Biometric Results for Word Recall

In the final set of analyses presented in Table (9) we examined the loss of information due to using only the single summary score of the Word Recall variable. The question we raise here is whether or not this single variable carries all the useful information available in these data. The first model (9a) is simply a repeat of the previous longitudinal model, this time without covariates. The second (9b) result is for the same model fitted only to the Immediate Word

recall, and the third result (9c) is the same model fitted to the Delayed Recall (after 5 minutes) of the same list of 10 words. The results for the means can be directly compared, and the Immediate Recall is easier than Delayed recall on average ($\beta_0=36$ vs 22) but it also declines more with age ($\beta_1=-4$ vs -2). The corresponding model for the covariances yields slightly different estimates but, most importantly, the pattern of biometric results is virtually identical across both variables (and as described above). The fourth model (9d) uses both sets of data on the two memory scores and gives estimates of the parameters of a common factor of episodic memory. The primary constraint is that this single factor accounts for the phenotypic covariation within a time, but the additional constraint is that this single factor must account for all of the useful information about the biometric structure of the twin data (as in McArdle & Goldsmith, 1990). This single latent common factor based on two indicators was initially fit within each wave, then across waves, and then a biometric structure was added. The primary result was that a single factor is a reasonable fit (with relative loadings of $\Lambda=1$ and 0.85), including one version requiring equal loadings for both memory variables and, once again, the relative biometric estimates were not substantially altered, and little information was lost due to our use of the summary Word Recall score.

DISCUSSION

Our key findings suggest that the longitudinal approach to biometric analysis can uncover the sources of change in cognitive functioning. The initial analysis showed a substantial impact of additive genetic sources on Educational Attainment, and these biometric results seem comparable to those found by others (e.g., Baker et al, 1995; Reynolds, et al, 2001). The observed changes of Word Recall showed approximately linear decline over this older age range. The cross-sectional (univariate) analyses did not yield high heritabilities (similar to Brandt et al, 1993), but it is possible that the sample loss due to attrition (death and dysfunction) led to initial underestimation of the genetic variance. In the longitudinal models presented here, the phenotypic age changes were found to be remarkably linear with the age-at-testing, but there were positive individual differences in initial level (at age 70 here) due to Retest (having had the test before) and Educational level. These same effects were not found for the age slopes, and this could be due to many different reasons. The subsequent biometric modeling of these variance components suggested that, while the Word Recall means decrease over age, the observed variance, additive genetic variance, and the independent environmental *increased* over age, producing dramatically decreasing shared family influences and increasing estimates of heritability in the later ages.

The substantive issues surrounding changes in genetic variance over age are not new (e.g., Wilson, 1978; Pedersen & Harris, 1991; McGue et al, 1993; Plomin et al, 1994; McArdle et al, 1998). However, any increases in genetic variance, such as those found here, require further elaboration. In this analysis, the males of the NAS-NRC Twin Registry were measured well past the ages of reproduction, so simple genetic selection up to the age of reproduction is not a reasonable explanation for the memory decline phenomena displayed in Figure (7). However, one important point is that in these models of growth and decline, the sources of the variance in the individual declines or losses (i.e., slopes) were not necessarily initiated in the adulthood range studied here but are considered to exist at all ages of measurement. If we were to extrapolate to earlier ages, these increasing genetic effects could be considered remnants of a genetic selection process that started much earlier in development and is only clearly expressed at the later ages (e.g., see Waddington, 1962). One interesting aspect of the final common factor model (9d) is the possibility that within time state variation, typically compounded with random measurement in latent curve models, can be isolated (i.e., correlation within time across both variables but not over time; McArdle & Woodcock, 1997). This would allow a potentially useful representation of stochastic aspects of within time measurement (Finch & Kirkwood, 2000), and these were not considered here. Nevertheless, these results do suggest that the

important variation in memory loss in the older adult age range is genetic in origin, and this has implications for further aging research (e.g., Whalley, 2001; Plassman et al, 2006).

These longitudinal biometric analyses also demonstrate functionally useful opportunities in the methodological integration of developmental, genetic, and statistical concerns. This mixed-effects model can be extended for use with categorical outcomes (see McArdle & Prescott, 2005), and this is important in the context of longitudinal studies. This is important because the NAS-NRC Twins Study was primarily designed as a longitudinal study of dementia. If individuals met pre-set criteria for cognitive impairment on a two-phase telephone screening protocol, they then were assessed further by a trained clinician and psychometrician who conducted a comprehensive in-person dementia evaluation. Onset of dementia was the age at which the subject unambiguously met the DSM-III-R criteria for dementia. If either twin had a diagnosis of dementia, this family was not followed up in subsequent waves of cognitive screening. However, it is possible that a small number of individuals with mild dementia were included in the present analyses if they crossed the threshold for dementia during the interval between two waves of cognitive screening.

This leads to at least two kinds of categorical models of interest. The first is a mixed-effects model formed from a survival-frailty model based on attrition from the study due to death or disease. Studies combining latent growth and survival models (e.g., McArdle et al, 2005) have shown standard software can be used for such analyses (i.e., NLMIXED, winBugs), and this includes multiple variable dynamic models (e.g., Mplus). The second is a mixed-effects model formed from a survival-frailty basis including probabilistic transitions to dementia (e.g., Meyer & Breitner, 1998; Ripatti, Gatz, Pedersen & Palmgren, 2003). Using the NAS-NRC data in this way we can separate out individuals with patterns of normative decline from patterns of pathological decline. This will allow us to examine the leading indicators of these transitions to dementia, such as early and mid-life activities and medical conditions, or the latent level and latent slope of the Word Recall task itself. Obviously, there are many new opportunities to merge novel and advanced modeling techniques with carefully collected large-scale longitudinal data sets to understand key aspects cognitive aging.

Acknowledgments

This research was supported by a grant to the first author from the *National Institute on Aging* (AG-07137) and to the second author (AG-008549).

REFERENCES

- Baker LA, Treloar SA, Reynolds CA, Heath AC, Martin NG. Genetics of educational attainment in Australian twins: Sex differences and secular changes. *Behav Genet* 1996;26:89–102. [PubMed: 8639155]
- Baltes, PB.; Nesselroade, JR. History and rationale of longitudinal research. In: Nesselroade, JR.; Baltes, PB., editors. *Longitudinal Research in the Study of Behavior and Development*. Academic Press; New York: 1979.
- Blum JE, Jarvik LF. Intellectual performance of octogenarians as a function of education and initial ability. *Human Development* 1974;17:364–375. [PubMed: 4468927]
- Brandt J, Welsh KA, Breitner JCS, Folstein MF, Helms M, Christian JC. Heredity influences on cognitive functioning of older men. *Archives of Neurology* 1993;50:599–603. [PubMed: 8503796]
- Brandt J, Spencer M, Folstein M. The Telephone Interview for Cognitive Status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology* 1988;1:111–117.
- Bialystok, E.; Craik, FIM., editors. *Lifespan Cognition: Mechanisms of Change*. Oxford University Press; New York: 2006.
- Brown, J., editor. *Recall and Recognition*. Wiley; New York: 1976.

- Christensen K, Frederiksen H, Vaupel JW, McGue M. Age trajectories of genetic variance in physical functioning: a longitudinal study of Danish twins aged 70 years and older. *Behav Genet* 2003;33:125–136. [PubMed: 14574147]
- Eaves LJ, Long J, Heath AC. A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behavior Genetics* 1986;16:143–162. [PubMed: 3707482]
- Finch, C.; Kirkwood, TBL. *Chance, Development and Aging*. Oxford; New York: 2000.
- Finkel D, Pedersen NL, Reynolds CA, Berg S, de Faire U, Svartengren M. Genetic and environmental influences on decline in biobehavioral markers of aging. *Behavior Genetics. Special Issue on Aging* 2003;33(2):107–123.
- Finkel D, Reynolds CA, McArdle JJ, Gatz M, Pedersen NL. Latent growth curve analyses of accelerating decline in cognitive abilities in late adulthood. *Developmental Psychology* 2003;39(3):535–550. [PubMed: 12760521]
- Finkel D, Reynolds CA, McArdle JJ, Pedersen NL. The longitudinal relationship between processing speed and cognitive ability: Genetic and environmental influences. *Behavior Genetics* 2005;35(5):535–550. [PubMed: 16184483]
- Frederiksen H, Gaist D, Bathum L, Andersen K, McGue M, Vaupel JW, Christensen K. Angiotensin I-converting enzyme (ACE) gene polymorphism in relation to physical performance, cognition and survival—a follow-up study of elderly Danish twins. *Ann Epidemiol* 2003;13:57–65. [PubMed: 12547486]
- Ghisletta P, McArdle JJ, Lindenberger U. Longitudinal cognition-survival relations in old and very old age: 13-Year Data from the Berlin Aging Study. *European Psychologist* 2006;11(3):204–223.
- Hewitt JK, Eaves LJ, Neale MC, Meyer JM. Resolving causing of developmental continuity or tracking: I. Longitudinal twin studies during growth. *Behavior Genetics* 1988;18:133–151. [PubMed: 3377729]
- Hultsch, DF.; Hertzog, C.; Dixon, R.; Small, BJ. *Memory Change in the Aged*. Cambridge University Press; Cambridge, UK: 1998.
- Jablon S, Neel JV, Gershowitz H, Atkinson GF. The NAS-NRC twin panel: methods of construction of the panel, zygosity diagnosis and proposed use. *American Journal of Human Genetics* 1967;19:133–161. [PubMed: 6067093]
- Johansson B, Hofer SM, Allaire JC, Maldonado-Molina MM, Piccinin AM, Berg S, Pedersen NL, McClearn GE. Change in cognitive capabilities in the oldest old: the effects of proximity to death in genetically related individuals over a 6-year period. *Psychology & Aging* 2004;19:145–156. [PubMed: 15065938]
- Katzman R, Aronson M, Fuld P, Kawas C, Brown T, Morgenstern H, et al. Development of dementing illnesses in an 80 year old volunteer cohort. *Annals of Neurology* 1989;25:317–324. [PubMed: 2712531]
- Loehlin JC. Partitioning environmental and genetic contributions to behavioral development. *American Psychologist* 1989;(10):1285–1292. [PubMed: 2679255]
- Loehlin JC, Horn JM, Willerman L. Modeling IQ change: Evidence from the Texas Adoption Project. *Child Development* 1989;(60):993–1004. [PubMed: 2758892]
- McArdle JJ. Latent variable growth within behavior genetic models. *Behavior Genetics* 1986;16(1):163–200. [PubMed: 3707483]
- McArdle JJ. Latent curve analysis of longitudinal twin data using a mixed-effects biometric approach. *Twin Research and Human Genetics* 2006;9(3):343–359. [PubMed: 16790145]
- McArdle JJ, Fisher GG, Kadlec KM. Latent Variable Analysis of Age Trends in Tests of Cognitive Ability in the Health and Retirement Survey, 1992-2004. *Psychology and Aging* 2007;22(3):525–545. [PubMed: 17874952]
- McArdle JJ, Goldsmith HH. Some alternative structural equation models for multivariate biometric analyses. *Behavior Genetics* 1990;20(5):569–608. [PubMed: 2288547]
- McArdle JJ, Hamagami F. Structural equation models for evaluating dynamic concepts within longitudinal twin analyses. *Behavior Genetics* 2003;33(3):137–159. [PubMed: 14574148]
- McArdle JJ, Prescott CA. Mixed-effects variance components models in the analysis of biometric and family data. *Behavior Genetics* 2005;34(5):631–652. [PubMed: 16184490]

- McArdle JJ, Prescott CA, Hamagami F, Horn JL. A contemporary method for developmental-genetic analyses of age changes in intellectual abilities. *Developmental Neuropsychology* 1998;14(1):69–114.
- McArdle JJ, Small BJ, Backman L, Fratiglioni L. Longitudinal models of growth and survival applied to the early detection of Alzheimer's Disease. *Journal of Geriatric Psychiatry and Neurology* 2005;18(4):234–241. [PubMed: 16306246]
- McArdle JJ, Woodcock JR. Expanding test-retest designs to include developmental time-lag components. *Psychological Methods* 1997;2(4):403–435.
- McClearn GE, Johansson B, Berg S, Pedersen NL, Ahern F, Petrill SA, et al. Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science* 1997;276:1560–1563. [PubMed: 9171059]
- McGue, M.; Bouchard, T.J.; Iacono, W.G.; Lykken, D.T. Behavioral genetics of cognitive ability: a life-span perspective. In: Plomin, R.; McClearn, G.E., editors. *Nature, nurture, and psychology*. American Psychological Association; Washington, DC: 1993. p. 59–76.
- McGue M, Christensen K. The heritability of level and rate-of-change in cognitive functioning in Danish twins aged 70 years and older. *Exp Aging Res* 2002;28:435–451. [PubMed: 12227922]
- Neale MC, McArdle JJ. The analysis of assortative mating: A LISREL model. *Behavioral Genetics* 1990;20(2):287–296.
- Neale MC, McArdle JJ. Structured latent growth curves for twin data. *Twin Research* 2000;3:1–13.
- Page WF, National Academy of Sciences-National Research Council. The NAS-NRC Twin Registry of WWII military veteran twins. *Twin Research* 2002;5:493–496. [PubMed: 12537883]
- Page WF. The NAS-NRC Twin Registry of WWII Military Veteran Twins. *Twin Research and Human Genetics* 2006;9:985–987. [PubMed: 17254441]
- Pedersen, N.L.; Harris, J.H. Developmental behavioral genetics and successful aging. In: Baltes, P.B.; Baltes, M.M., editors. *Successful aging: Perspectives from the behavioral sciences*. Cambridge University Press; Cambridge: 1991. p. 359–380.
- Pedersen NL, Plomin R, Nesselroade JR, McClearn GE. A quantitative genetic analysis of cognitive abilities during the second half of the life-span. *Psychological Science* 1992;3:346–353.
- Pedersen NL, Ripatti S, Berg S, Reynolds C, Hofer SM, Finkel D, et al. The influence of mortality on twin models of change: Addressing missingness through multiple imputation. *Behavior Genetics* 2003;33(2):161–169. [PubMed: 14574149]
- Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology* 1999;56:303–308. [PubMed: 10190820]
- Plomin R, Pedersen NL, Lichtenstein P, McClearn GE. Variability and stability in cognitive abilities are largely genetic later in life. *Behavior Genetics* 1994;24:207–216. [PubMed: 7945151]
- Plassman BL, Newman TT, Welsh KA, Helms M, Breitner JCS. Properties of the Telephone Interview for Cognitive Status. Application in epidemiological and longitudinal studies. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology* 1994;7:235–241.
- Plassman BL, Steffens DC, Burke JR, Welsh-Bohmer KA, Newman TN, Drosdick D, Helms MJ, Potter GG, Breitner JCS. Duke Twins Study of Memory in Aging in the NAS-NRC Twin Registry. *Twin Research and Human Genetics* 2006;9:950–957. [PubMed: 17254435]
- Plassman BL, Welsh KA, Breitner JCS, Brandt J, Helms M, Page WF. Intelligence and education as predictors of cognitive state in late life: A 50 year follow-up. *Neurology* 1995;45:1446–1450. [PubMed: 7644038]
- Potter GG, Plassman BL, Helms MJ, Foster SM, Edwards NW. Occupational characteristics and cognitive performance among elderly male twins. *Neurology* 2006;67:1377–1382. [PubMed: 17060563]
- Potter GG, Plassman BL, Helms MJ, Steffens DC, Welsh-Bohmer KA. Age effects of coronary artery bypass graft on cognitive status change among elderly male twins. *Neurology* 2004;63:2245–2249. [PubMed: 15623681]
- Reed T, Plassman BL, Tanner CM, Dick DM, Rinehart SA, Nichols WC. Verification of self-report of zygosity determined via DNA testing in a subset of the NAS-NRC Twin Registry 40 years later. *Twin Research and Human Genetics* 2005;8:362–367. [PubMed: 16176721]

- Reynolds CA, Finkel D, Gatz M, Pedersen NL. Sources of influence on rate of cognitive change over time in Swedish twins: An application of latent growth models. *Experimental Aging Research* 2002;28(4):407–433. [PubMed: 12227921]
- Reynolds CA, Finkel D, McArdle JJ, Gatz M, Berg S, Pedersen NL. Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. *Developmental Psychology* 2005;41(1):3–16. [PubMed: 15656733]
- Reynolds CA, Gatz M, Pedersen NL. Individual variation for cognitive decline: Quantitative methods for describing patterns of change. *Psychology & Aging* 2002;17(2):271–287. [PubMed: 12061412]
- Reynolds CA, Jansson M, Gatz M, Pedersen NL. Longitudinal change in memory performance associated with HTR2A polymorphism. *Neurobiology of Aging*. 2007 in press.
- Steuer J, LaRue A, Blum J, Jarvik LF. Critical loss in the eighth and ninth decades. *Journal of Gerontology* 1981;36:211–213. [PubMed: 7193706]
- Vandenberg SG, Falkner F. Heredity factors in human growth. *Human Biology* 1965;37:357–365.
- Waddington, CH. *New patterns in genetics and development*. Columbia University Press; New York: 1962.
- Wilson R. Continuity and change in cognitive ability profile. *Behavior Genetics* 1986;16:45–60. [PubMed: 3707487]
- Wetherell JL, Reynolds CA, Gatz M, Pedersen NL. Anxiety, cognitive performance, and cognitive decline in normal aging. *Journals of Gerontology: Series B: Psychological Sciences & Social Sciences* 2002;57B(3):P246–P255.
- Welsh KA, Breitner JCS, Magruder-Habib KM. Detection of dementia in the elderly using telephone screening of cognitive status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology* 1993;6:103–110.
- Whalley, L. *The Ageing Brain*. Pheonix; London: 2001.
- Xiong GL, Plassman BL, Helms MJ, Steffens DC. Vascular risk factors and cognitive decline among elderly male twins. *Neurology* 2006;67:1586–1591. [PubMed: 17101888]

Figure [1].
 Descriptive Statistics at Wave[1]

a

NAS—NRC Longitudinal Twin=1 Data Up to Four Waves

b

NAS—NRC Longitudinal Twin=1 Data Up to Four Waves

Figure [2].

- a: Plotting Words over 4 Waves (5% random sample, one person per family)
- b: Plotting Words over Multiple Ages (5% random sample, one person per family)

Figure [3].
A path diagram of an LGM for longitudinal twin data

Figure [4].
Results for Educational Attainment from all twin data at Wave[1] ($N=6,488$, $D=11,796$)

Figure [5]. Results for All Words Recalled from all twin data at Wave [1] ($N=6,440$, $D=11,597$)

Figure [6].

Summary of longitudinal results from the *Linear* decline curve for Words Recalled over all ages from four waves (Covariates not drawn; $N=6,560$ pairs; $D=31,629$)

Figure [7].
Expected Deviations over Age from Linear Decline Latent Curve Model

Table 1
Summary of Available Longitudinal Data for the NAS-NRC Participants at Up to Four Waves Testing

Testing Label	Frequency of People	% of People	Frequency of Tests	% of All Tests	Formal Education In Years (mean)	Age at First Test (mean)	Word Recall Score at First Test (mean)
(a) Persons Measured at all Occasions							
1234	3,948	29.2%	15,792	45.3%	14.0	65.8	32.8
(b) Persons Measured on Three Occasions							
1230	2,159	16.0	6,477	18.6	12.9	66.5	29.3
1204	254	1.9	762	2.2	12.9	66.1	30.3
0234	137	1.0	411	1.2	?	68.8	28.8
1034	88	<1	264	<1	13.4	66.0	30.9
(c) Persons Measured on Two Occasions							
1200	3,882	28.7	7,764	22.3	12.5	66.8	27.6
1030	171	1.3	342	<1	12.5	66.4	26.9
0230	115	<1	230	<1	?	69.3	24.5
1004	35	<1	70	<1	12.4	66.4	25.0
0204	19	<1	38	<1	?	69.6	23.7
0034	3	<1	6	<1	?	72.6	18.3
(d) Persons Measured on One Occasion							
1000	2,172	16.1	2,172	6.2	12.3	66.9	25.6
0200	539	4.0	536	1.5	?	69.8	21.6
0030	10	<1	10	<1	?	71.9	20.5
(e) Total Across All Persons							
	13,552	100%	34,874	100%	13.0	66.6	28.9%

Notes:
Each numeral in the testing label indicates the number of the wave(s) completed by the subgroup.
For example, subgroup with Testing label = 1234 completed all four waves of cognitive screening.
Frequency of tests was determined by multiplying the number of waves of cognitive screening completed by the 'Frequency of People'.
Word Recall Score is the average of the scores for immediate plus delayed recall multiplied by 10.

The “?” indicates the data were not available for most persons in the cell.

Table [2]
Summary Statistics for the Total Words Recalled at Four Waves of the data for the MZ-DZ pairs

(a) Available Data for Mono-Zygotic Twins						
MZ Wave	MZ families	MZ data points	Mean	Common Variance	Residual Variance	Intraclass Correlation
Age[1]	3,010	5,535	66.4	8.09	0.13	0.98
Education	3,008	5,539	13.3	7.56	2.27	0.770
Words[1]	2,987	5,529	30.0	67.8	135.0	0.335
Words[2]	2,686	4,735	30.2	79.7	159.6	0.333
Words[3]	1,812	3,146	32.4	100.1	178.3	0.361
Words[4]	1,428	2,253	28.4	88.3	148.3	0.373

(b) Available Data for Di-Zygotic Twins						
DZ Wave	DZ families	DZ data points	Mean	Common Var.	Residual Var.	Intraclass Correlation
Age[1]	3,483	6,256	66.5	8.26	0.14	0.97
Education	3,481	6,257	12.9	5.77	4.43	0.566
Words[1]	3,453	6,068	28.6	41.8	150.0	0.218
Words[2]	3,007	5,048	29.3	48.6	186.8	0.207
Words[3]	1,883	3,005	31.8	58.0	211.8	0.215
Words[4]	1,364	2,000	28.4	49.5	199.8	0.199

Notes:
Age[t]≥59-85; Wave[t]=1-4 Age [1] – mean age at Wave 1.
Education – self-reported in years.
Words [t] is the Episodic Memory Score for waves [1]–[4].
The sample sizes (*N families*) and the amount of available score data (*D – data points*) at each wave of testing are presented in the next two columns.
Mean -overall mean within groups.
Common variance - variance that is common to the pair (σ_0^2).
Residual variance – variance that is unique to each person (σ_U^2).
The likelihood (L^2) represents the loss of information based on assuming these terms apply to all members of a pair.
Intra-class correlation coefficient (η^2) calculated within pairs.

Table [3]

Cross-Sectional Biometric Models for Educational Attainment

<i>Biometric Parameters</i>	<i>(a) Initial Biometric Decomposition</i>		<i>(b) Biometric Decomposition with Age Covariate</i>	
	<i>Full MLE (z-val)</i>	<i>Reduced MLE (z-val)</i>	<i>Full MLE (z-val)</i>	<i>Reduced MLE (z-val)</i>
<i>(a) Fixed Effects</i>				
Intercept @ Age 70 β_0	13.04 (28)	13.03 (30)	12.85 (15)	12.84 (15)
Slope/Decade β_1			-0.54 (-4.3)	-0.54 (-4.6)
<i>(b) Random Effects</i>				
Additive σ_a^2	4.29 (16)	7.66 (52)	4.30 (16)	7.63 (52)
Shared σ_s^2	3.52 (14)	=0	3.48 (13)	=0
Independent σ_i^2	2.27 (35)	2.12 (38)	2.27 (35)	2.13 (38)
<i>(c) Variance Percentage</i>				
Additive η_a^2	42.5%	78.3%	42.8%	78.2%
Shared η_s^2	34.9%	=0%	34.6%	=0%
Independent η_i^2	22.5%	21.7%	22.6%	21.8
<i>(d) Goodness-of-Fit</i> <i>Biometric $\Delta L^2/\Delta df$</i> <i>No Shared $\Delta L^2/\Delta df$</i>	29 / 2	183 / 3 154 / 1	67 / 1	209 / 2 142 / 1

Note: $N=6,488$ pairs with Education data on $D=12,709$ persons.

Table [4]

Results of Univariate Twin Models for Word Recall at Each Wave

<i>Parameter</i>	<i>Maximum Likelihood Estimates</i>			
<i>(a) Fixed Effects</i>	Words[1]	Words[2]	Words[3]	Words[4]
Mean μ_0	29.3	29.7	32.1	28.4
<i>(b) Random Effects</i>				
Additive σ_a^2	43.8	54.2	63.1	94.4
Shared σ_s^2	21.6	24.1	32.5	0 ^{ns}
Independent σ_i^2	131.9	159.0	178.9	148.7
<i>(c) Variance Percentage</i>				
Additive η_a^2	22.2%	22.8%	23.0%	38.8%
Shared η_s^2	11.0%	10.2%	11.8%	0%
Independent η_i^2	66.8%	67.0%	65.2%	61.2%
<i>(d) Goodness-of-Fit</i>				
Biometric $\Delta L^2/\Delta df$	26 / 5	8 / 5	3 / 5	1 / 5
No Shared $\Delta L^2/\Delta df$	8 / 1	4 / 1	1 / 1	0 / 1

Note: For sample sizes, see Tables 1 and 2; “ns” indicates parameter with $z=\text{est}/\text{se}<3.0$.

Table [5]

Results Including Covariates in Univariate Twin Models for Word Recall

<i>Parameter</i>	<i>Maximum Likelihood Estimates</i>			
<i>(a) Fixed Effects</i>	Words[1]	Words[2]	Words[3]	Words[4]
Intercept@ Age 70 β_0	25.4	27.5	31.8	33.7
Age Slope β_1	-6.56	-8.90	-8.36	-9.92
Education Slope β_1	1.41	1.52	1.68	1.45
<i>(b) Random Effects</i>				
Additive σ_a^2	33.1	38.4	47.9	46.8
Shared σ_s^2	6.5 ^{ns}	7.8 ^{ns}	13.3 ^{ns}	13.5 ^{ns}
Independent σ_i^2	132.4	158.6	180.4	152.5
<i>(c) Variance Percentages</i>				
Additive η_a^2	19.2%	18.8%	19.8%	22.0%
Shared η_s^2	3.8%	3.8%	5.5%	6.4%
Independent η_i^2	77.0%	77.4%	74.7%	71.7%
<i>(d) Goodness-of-Fit</i>				
Covariate $\Delta L^2/\Delta df$	1445 / 2	6702 / 2	2698 / 2	1697 / 2
No Shared $\Delta L^2/\Delta df$	0 / 1	0 / 1	0 / 1	0 / 1

Note: For sample sizes, see Tables 1 and 2; "ns" indicates parameter with $z=\text{est}/\text{se}<3.0$.

Table [6]
Results of Alternative Phenotypic Linear Latent Curve Models using Total Word Recall Scores from NAS-NRC Longitudinal Individual Data

Parameter	(a) Linear Slope Model		(b) Linear Slope Model With Education		(c) Linear Slope Model with Education and Retest	
	Means	Means +Covars	Level Effects	+Slope Effects	Level Effects	+ Slope Effects
(a) Fixed Effects						
Intercept @ Age 70 β_0	29.2	29.2	27.2	27.4	25.7	26.1
Age Slope 1/Decade β_1	-2.94	-3.17	-3.34	-3.62	-5.53	-4.68
Education β_{0e}			1.59	1.61	1.57	1.49
Education on Age β_{1e}				0.18 ^{ns}		0.02 ^{ns}
Retest/Practice β_{0r}					2.79	2.31
Retest on Age Slope β_{1r}						-1.32 ^{ns}
Retest x Education β_{0i}						0.14 ^{ns}
Retest x Educ on Slope β_{1i}						0.11 ^{ns}
(b) Random Effects						
Unique σ_u^2	129.6	128.7	128.4	128.3	127.7	127.4
Intercept σ_0^2	99.9	101.7	77.1	77.4	75.7	76.1
Age Slope σ_1^2		4.9	4.8	4.7	5.9	5.7
Correlation ρ_{01}		0.89	0.83	0.85	0.75	0.77
(c) Variance Percentages @Age 80						
Unique η_u^2	56.5%	46.8%	47.9%	47.9%	47.8%	47.7%
Intercept η_0^2	43.5%	44.2%	34.8%	34.8%	31.7%	31.6%
Age Slope η_1^2		9.0%	7.8%	7.8%	8.4%	8.3%
Education η_e^2			9.5%	9.5%	9.4%	9.1%
Retest η_r^2					2.7%	2.7%
(d) Goodness-of-Fit						
L^2 / parms	3315 / 4	3261 / 6	2190 / 7	2185 / 8	2068 / 8	2054 / 12

Parameter	(a) Linear Slope Model		(b) Linear Slope Model With Education		(c) Linear Slope Model with Education and Retest	
	Means	Means +Covars	Level Effects	+Slope Effects	Level Effects	+ Slope Effects
(a) Fixed Effects						
Model $\Delta L^2/\Delta df$	194 / 1	54 / 2	1071 / 1	5 / 1	122 / 1	14 / 4

Notes:

These models were fitted using only one twin per family -- N=6,349 persons measured on up to T=4 occasions using D=16,447 data points.

For presentation here, all L^2 presented are reduced by 130000.

The full random intercept model has $L^2=3499$ with 3 parameters ($\mu\sigma^2=29.3$, $\sigma_{it}^2=131$, and $\sigma\sigma^2=100$) so the initial $\eta\sigma^2=43.3\%$.

The “ns” indicates parameter with $z_{est/se}<3.0$.

Table [7]

Results of Linear Growth Biometric Modeling using the Total Word Recall Scores from all NAS-NRC Twin Pairs at all Longitudinal Waves

<i>Parameter</i>	<i>(a) Overall Biometric Intercept and Linear Slope Model</i>		<i>(b) Reduced Biometric Intercept and Linear Slope Model</i>	
<i>(a) Fixed Effects</i>	<i>Estimate</i>	<i>Variance %</i>	<i>Estimate</i>	<i>Variance %</i>
Intercept @70 β_0	25.6	=	25.6	=
Age Slope/Decade β_1	-5.63	=	-5.63	=
Education β_e	1.51	=	1.51	=
Retest β_r	2.87	=	2.86	=
<i>(b) Random Effects and Variance Percentages</i>				
Unique Variance $\sigma_u^2 (\eta_u^2)$	123.9	58.3%	123.0	59.6 %
<i>Additive Genetic</i>				
<i>u0-Additive</i> $\sigma_{u0a}^2 (\eta_{0a}^2)$	28.1	50.7%	24.0	50.0%
<i>u1-Additive</i> $\sigma_{u1a}^2 (\eta_{1a}^2)$	9.5	>100%	0.27	100.0%
<i>c01-Additive</i> σ_{01a}^2	11.5	=	15.0	=
<i>Shared Environmental</i>				
<i>u0-Shared</i> $\sigma_{u0s}^2 (\eta_{0s}^2)$	26.2	7.1%	32.2	8.6%
<i>u1-Shared</i> $\sigma_{u1s}^2 (\eta_{1s}^2)$	-16.9	<0%	=0	0%
<i>c01-Shared</i> σ_{01s}^2	6.7	=	=0	=
<i>Independent Environmental</i>				
<i>u0-Independent</i> $\sigma_{u0i}^2 (\eta_{0i}^2)$	6.6	42.2%	6.7	41.4%
<i>u1-Independent</i> $\sigma_{u1i}^2 (\eta_{1i}^2)$	0.8	<0%	=0	0%
<i>c01-Independent</i> σ_{01i}^2	-1.1	=	=0	=
<i>(c) Goodness-of-Fit</i>				
<i>LRT / parms</i>	590/ 15		604/11	
<i>L² / df</i>	0 / 0		14 / 4	

Notes:

These models were fitted using both twins -- $N=6,486$ pairs, measured on up to $T=4$ occasions using $D=30,693$ data points.

The LRTs presented are reduced by 245000.

The "ns" indicates parameter with $z=\text{est}/\text{se}<3.0$.

Components calculated from Education ($m=1.2$, $sd=3.2$) and Retest ($m=0.6$, $sd=.5$).

Table [8]
Age-Related Biometric Parameters for Total Words Recalled from the Longitudinal Twin data (from Model 7a)

Estimated Parameters	(a) Phenotypic Statistics over Age			(b) Raw Score Deviation Over Age			(c) Percentage of Total Variance Over Age		
	Mean $\mu[t]$	Dev. $\sigma[t]$	Reliab. $\eta[t]^2$	Genet. $\sigma_g[t]$	Shared $\sigma_s[t]$	Indep. $\sigma_i[t]$	Genet. $\eta_g[t]^2$	Shared $\eta_s[t]^2$	Indep. $\eta_i[t]^2$
Age in Years									
60	29.3	7.4	.305	6.1	2.8	3.1	69.1	13.7	17.2
65	27.0	8.0	.340	5.8	2.6	4.9	52.4	10.3	37.3
70	24.7	8.8	.386	6.3	2.4	5.7	50.7	7.1	42.2
75	22.3	9.9	.441	7.5	2.0	6.1	57.5	4.5	37.2
80	20.0	11.1	.498	9.1	1.8	6.0	68.0	2.6	29.4
85	17.7	12.4	.552	11.0	1.3	5.5	79.1	1.2	19.7
90	15.4	13.7	.604	13.0	0.6	4.4	89.7	0.2	10.1

Table [9]
Results of Longitudinal Biometric Modeling using different versions of Word Recall Scores ($N=6,560$ pairs for $T=4$ and $D=31,629$)

Parameter	(a) SUM ALL WORDS		(b) IMMEDIATE RECALL		(a) DELAYED RECALL		(d) COMMON FACTOR	
	Est.	(z-Val)	Est.	(z-Val)	Est.	(z-Val)	Est.	(z-Val)
Fixed Effects								
Intercept @70 β_0	28.9	250	35.5	248	22.4	161	24.7	114
Age Slope/Decade β_1	-3.17	-18	-4.21	-21	-2.02	-10	-4.62	-17
Random Effects								
Unique Error σ_u^2	125.3	86	167.3	87	145.6	87	124.0 ($\lambda=85$)	86 (95)
Additive Genetic								
$u0$ -Additive σ_{u0a}^2	39.1	5	41.7	4	39.1	4	28.2	3
$u1$ -Additive σ_{u1a}^2	5.1 ^{ns}	0	12.2 ^{ns}	0	6.2 ^{ns}	0	9.3 ^{ns}	1
$c01$ -Additive σ_{01a}^2	12.3	2	13.7	2	14.1	2	12.0	1.8
Shared Environ.								
$u0$ -Shared σ_{u0s}^2	21.1	3	6.3	3	22.0	3	7.1 ^{ns}	1
$u1$ -Shared σ_{u1s}^2	3.5 ^{ns}	0	-4.1 ^{ns}	0	2.0 ^{ns}	0	2.1 ^{ns}	0
$c01$ -Shared σ_{01s}^2	-6 ^{ns}	0	-1.9 ^{ns}	0	-6 ^{ns}	0	-3.2 ^{ns}	-1
Independent Environ.								
$u0$ -Independent σ_{u0f}^2	24.9	9	21.3	9	35.9	10	26.1	9
$u1$ -Independent σ_{u1f}^2	-19.3	-3	-25.1	-3	-15.4	-2	-16.7	3
$c01$ -Independent σ_{01f}^2	8.3	4	9.0	4	3.6*	1	6.6	3
Goodness-of-Fit								
χ^2 / pams	10144 / 12		17313 / 12		14292 / 12		9709 / 13	

Notes:
These models were fitted using $N=6,486$ pairs, measured on up to $T=4$ occasions using $D=30,693$ data points.
The LRT 's presented are reduced by 245000, and final model by 2*245000.