Growing Old but Not Growing Apart: Twin Similarity in the Latter Half of the Lifespan

Matt McGue and
Department of Psychology/Elliott Hall, University of Minnesota, Minneapolis, MN 55455, USA.
Institute of Public Health, University of Southern Denmark, Odense, Denmark

Kaare Christensen
The Danish Twin Registry and The Danish Aging Research Center Institute of Public Health, University of Southern Denmark, Odense, Denmark

Matt McGue: mcgue001@umn.edu

Abstract

While a substantial amount of behavioral genetic research has helped to characterize developmental trends in twin similarity in early life, relatively little is known about changes in twin similarity with age in adulthood. We investigated age moderation of twin similarity for a composite measure of cognitive ability, depression symptomatology and hand grip strength in a cross-sectional sample of 2,332 like-sex pairs of Danish twins age 46–96 years. All three outcomes were strongly correlated with age, indicating that the three phenotypes analyzed are not developmentally static. Nonetheless, in moderated regression analysis we found no evidence of declining twin similarity for any of the three outcomes in either zygosity group. Moreover, biometric analysis of the twin data revealed minimal differences in heritability estimates across the age range sampled. While small sample size limits our ability to draw firm conclusions at very advanced ages, these findings call into question the hypothesis that the cumulative impact of life experiences diminishes twin similarity at least through age 80. We hypothesize that twins are able to maintain similarity over extended periods of time because in part they are able to construct similar environments that reinforce that similarity.

Keywords

Aging; Depression; Cognitive functioning; Grip strength; Age differences in heritability; Twin correlation

Introduction

Whether twins become more or less similar with age is a question nearly as old as the Nature–Nurture debate. Galton (1875) may have been the first to recognize the importance of investigating developmental shifts in twin similarity. In the absence of a reliable method for differentiating monozygotic (MZ) twins from dizygotic (DZ) twins, Galton reasoned that developmental changes in twin similarity could provide insights into the relative contribution of heritable and environmental influences on behavior. Specifically, anecdotal observations suggesting that twins did not become behaviorally less similar as the twins

grew older, led Galton to his famous conclusion that, "nature prevails enormously over nurture" (p. 576).

It took more than a century before Galton's developmental hypothesis received the thorough and quantitative analysis it deserved. In a meta-analysis of more than 100 twin studies published between 1967 and 1985, McCartney et al. (1990) concluded that, "as twins grow up, they grow apart". (p. 236). For both MZ and DZ twins and for measures of personality as well as cognitive ability, the general trend was for the twin correlation to decline with age. Although rigorous, the McCartney et al. (1990) meta-analysis was not without its limitations. Of greatest relevance to the present study, the overwhelming bulk of the studies reviewed involved twins in childhood or early adolescence. That is, the declining twin resemblance McCartney et al. (1990) observed may have owed largely to the now well-known phenomenon that for most psychological traits the magnitude of shared environmental influence declines from childhood through early adulthood (Eaves et al. 1997; Koenig et al. 2008; McGue et al. 1993). Consistent with this interpretation, in a recent meta-analysis of behavioral genetic studies spanning essentially the same ages as those included in the earlier meta-analysis by McCartney et al. (1990), Bergen et al. (2007) concluded that for most of the behavioral traits studied heritability increased from adolescence through early adulthood.

Nearly absent from the existing research literature, however, is a consideration of changes in twin similarity and biometric variance component estimates during adulthood, and especially in late-life. A notable exception is the recent chapter by Finkel and Reynolds (2010), which provides a comprehensive review of evidence for genetic contributions to late-life change in cognitive function. Aggregating findings from studies of general cognitive ability in adult twins led these researchers to conclude, "both cross-sectional and longitudinal studies converges on the conclusion that heritability for general cognitive ability increases from young adulthood, plateaus in adulthood, and decreases late in life" (p. 102). There are several reasons to expect that twin similarity might change with age in adulthood. First, twin similarity might decline as a consequence of the cumulative impact of non-shared environmental influences. Even if not independent, the experiences of two members of a twin pair certainly diverge throughout life and especially once they attain adulthood and establish separate careers, families and homes. Consequently, following Galton's original logic, we might expect that the cumulative impact of their divergent lifestyles would lead twins to become less similar as they age. Alternatively, although not consistent with the conclusion by Finkel and Reynolds (2010) concerning general cognitive ability, twin similarity might increase with age. For example, at least one evolutionary model of biological aging, the mutation accumulation model (Hamilton 1966), posits that genetic variance increases with age because deleterious late-acting mutations will not have been exposed to the homogenizing effects of natural selection. Thus under this model we might expect heritability to increase with age.

Research investigating heritability changes in adulthood for behavioral traits has focused largely on cognitive ability, and generally supports the conclusion reached by Finkel and Reynolds (2010). The pioneering work in this area has been undertaken by Finkel and her colleagues (1995, 1998). In a joint cross-sectional analysis of data from the Swedish Adoption/Twin Study of Aging (SATSA) and the Minnesota Twin Study of Adult Development and Aging (MTSADA), Finkel (Finkel et al. 1995) reported that the heritability of general cognitive ability declined from 81 % in early adulthood/middle age to 54 % in late-adulthood (which was defined as being older than 65) in the Swedish sample. A similar decline in heritability was not, however, observed in the Minnesota sample. In a subsequent cross-sequential analysis that made use of three waves of cognitive ability assessment in SATSA, Finkel et al. (1998) confirmed an age-related decrease in the

heritability of general cognitive ability after age 70, as measured by the first principal component of a battery of 13 cognitive ability tests. An analysis of unstandardized variance components indicated that the observed decline in the heritability was the result of declines in genetic variance coupled with increases in non-shared environmental variance. Nonetheless, several of the individual tests used in forming the overall cognitive composite showed a pattern of increasing heritability with age (e.g., picture memory and figure identification), and the authors concluded that, "no general pattern of change in components of variance was found" (p. 1406).

A subsequent cross-sequential study of SATSA cognitive ability data made use of a fourth wave of assessment and further extended previous analysis of the SATSA data by considering non-linear as well as linear patterns of change, finding evidence for both (Reynolds et al. 2005). As before, a pattern of declining genetic influence was observed for the first principal component of the cognitive ability data. The estimated heritability of general cognitive ability was approximately 80 % for age 50–70 but 64 % at age 80. The decline in heritability was due to both an increase in the unstandardized non-shared environment component of variance and a decrease in the genetic component of variance in the oldest as compared to the other age groups. Alternatively, although the estimate of the non-shared environmental component of variance increased with age for nearly all of the 13 specific cognitive measures, the estimate of genetic variance was stable for some measures (e.g., synonyms), increasing for others (e.g., information, both memory measures), and markedly decreasing for others (e.g., block design, figure logic).

In a 12-year longitudinal study that is notable in terms of its large sample of twins (>6,000 pairs), McArdle and Plassman (2009) used twins initially aged 59–75 from the NAS–NRC Twin Registry of WWII veterans to investigate genetic effects on episodic memory. Based on a non-linear two-part spline curve model that fit the combined longitudinal and cross-sectional data well, they concluded that genetic effects on episodic memory increased through age 74 (consistent with the SATSA finding that genetic variance estimates for memory measures increased with age) and slowly declined thereafter.

Depression, the second phenotype investigated in the present study, is the only other phenotype for which there are multiple twin studies investigating age moderation of heritable effects in late-life. The largest twin study to investigate age moderation effects is the study by Johnson et al. (2002), who found no evidence of age moderation of heritable effects on a quantitative measure of depression symptomatology in a sample of 2,169 Danish twins aged 46–95 (a sample that overlaps extensively with the sample used here). Alternatively, in a sample of 481 reared-together and reared-apart twins from SATSA, Gatz et al. (1992) estimated the heritability of a quantitative measure of depression symptomatology to be 18 % for twins aged 60 years or older but only 3 % for twins younger than age 60, a difference that was statistically significant. Similarly, in a longitudinal study of 187 twin pairs initially aged 59–70, Carmelli et al. (2000) reported that the heritability of a quantitative index of depression symptomatology increased from 25 % at baseline to 55 % at 10-year follow-up. Alternatively, rather than an increase in heritability with age, Whitfield et al. (2008) reported that the heritability of liability to a categorical indicator of depression decreased from 28 % for twins 49 years and younger to 6 % for twins 50 years and older in a sample of 212 twin pairs. Finally, Neiss and Almeida (2004) reported evidence of both increases and decreases in heritable influences on different measures of psychological distress from a single sample of 210 pairs of twins age 25–74 years from the National Survey of Midlife in the United States Survey (MIDUS). Specifically, they reported that the heritability of psychological distress increased with age when the reporting interval was the past month but decreased with age when it was reported over the past 24 h.

Only for grip strength, the third phenotype investigated here, is the literature on age moderation of heritable effects consistent, even if based on only two studies. Using SATSA data, Finkel et al. (2003) reported that genetic effects on hand grip strength were relatively constant from age 50 to 90. Similarly, using data from a sample of 1,757 Danish twin pairs (which overlaps extensively with the sample used here), Frederiksen et al. (2002) found no evidence of age moderation of heritable effects on grip strength.

The existing literature on changes in the magnitude of genetic influences on behavioral traits in adulthood is thus sparse and in most cases yields a complex set of findings that defies easy summary. One limitation of much of the existing research is that, with the exception of the McArdle and Plassman (2009) study and the Danish cross-sectional studies, the twin samples are small, typically several hundred. This is to be expected in aging research, where it can be difficult to identify and ascertain large samples of intact twin pairs. Nonetheless, small sample sizes raise several potential problems for investigating age moderation of heritable effects. Most obviously, the power to test for age-moderated effects will be low in small samples, likely resulting in increased rates of false negative findings and failures to replicate. Alternatively, variance component estimates might be unduly susceptible to the influence of extreme values when there are small samples, especially at the most advanced ages, in which case a small sample might, counter-intuitively, increase the likelihood of a false positive finding.

The goal of the current paper is to use data from a large cross-sectional sample of twins to determine whether there is a coherent pattern of adult changes in twin similarity and heritability much like the coherent set of behavioral genetic findings that has emerged on the impact of the transition from adolescence to early adulthood. Specifically, we make use of a cross-sectional sample of more than 2,000 pairs of like-sex Danish twins, aged 46–96, who have all been assessed on the same measures of general cognitive ability, depression symptomatology, and hand grip strength. Using both moderated regression and biometric analysis, we sought to determine whether there are consistent changes in both twin similarity and the magnitude of genetic effects in the latter half of the lifespan. We have previously reported age comparisons of twin similarity for the depression (Johnson et al. 2002) and grip strength (Frederiksen et al. 2002) measures. Neither study found evidence of differences in twin correlation as a function of age group. In both cases, however, age was treated as a discrete entity and only a single approach to age moderation was investigated. Here we increase the sensitivity of the analysis by treating age as a continuous quantitative variable, considering both linear and non-linear effects, and using both regression and biometric modeling analysis to test for age moderation effects. As a consequence, the analysis reported here should provide one of the most sensitive tests of the age moderation question yet undertaken. No previous cross-sectional analysis of twin similarity for the Danish cognitive data has been reported.

Method

Sample

The sample was drawn from the more than 9,000 participants in two studies undertaken under the auspices of the Danish Twin Registry (Skytthe et al. 2002). The first is the Middle Age Danish Twin (MADT) study, which targeted a sample of 120 twin pairs in each of the 22 consecutive birth cohorts from 1931 to 1952 (Gaist et al. 2000). Of the 5,280 individual twins in the MADT sampling framework, 90 died prior to the time the survey was undertaken and 4,314 (83.1 %) of the 5,190 surviving twins completed an in-person intake assessment in late 1998 or early 1999, at which time the twins varied in age from 46 to 68 years. The second study is the Longitudinal Study of Aging Danish Twins (LSADT, Christensen et al. 1999). LSADT began in 1995 with the assessment of members of like-sex

twin pairs born in Denmark prior to 1920 (i.e., at least 75 years old). The surviving members of the initial cohort were followed up every 2 years in 1997, 1999, 2001, 2003, and 2005. Additional cohorts were added in 1997, 1999, and 2001 and the minimum age requirement was progressively reduced to age 70. Of a total of 6,542 eligible twins, 4,731 (72.3 %) completed an intake LSADT assessment. In order to ensure comparability of data across the two surveys, only the initial assessment from the LSADT is used in the current report. Consequently, all of our analyses are cross-sectional.

Although we make use of data from all available participants to describe age differences in the key outcomes we investigate, our primary focus is on cross-sectional differences in twin resemblance. Table 1 provides a breakdown of the number of MZ and like-sex DZ twins by age group in the two studies. In total, there were 2,332 pairs of twins available for analysis, although as described below the sample sizes vary a bit from measure to measure due to missing data. Of note is the fact that LSADT recruited twins regardless of whether an individual's co-twin was alive. As a consequence, the number of intact twin pairs available for analysis is much less than half the number of individual participants.

Assessment

Assessment procedure—Both studies used a very similar ascertainment and assessment protocol. Living twins were identified through the Danish Central Persons Register and sent a letter inviting them to participate in a survey. In the vast majority of cases, the survey took place in the participant's home and was administered by one of approximately 100 interviewers from the Danish National Institute of Social Research, which has extensive experience in undertaking surveys with elderly Danes (Kjøller 1996; Platz 1989, 1990). Interviewers completed a detailed training program 2 months prior to survey administration and were closely monitored during the survey period.

Depression symptoms—In both surveys, depression symptomatology was assessed using an adaptation of the depression section of the Cambridge Mental Disorders of the Elderly Examination (CAMDEX, Roth et al. 1986). McGue and Christensen (1997) factor analyzed the CAMDEX depression items and identified two factor scales (a nine-item Affective and an eight-item Somatic scale), aswell as a 17-item total score. It is the total score that is used here. The total depression scale has high internal consistency reliability (.83) and is moderately stable over a 2-year period (.64). Among the 9,045 total participants in the two surveys, 8,736 (97 %) individuals and 2,307 twin pairs completed an intake assessment of depression. The major reason for missing data is proxy response in the LSADT survey. To minimize skewness, the depression symptom score was log transformed prior to all analysis reported here.

Cognitive composite—Cognitive functioning was assessed as a composite of five individual tests selected to be sensitive to normative age changes but could be adapted for brief and reliable administration by lay interviewers. The specific tasks included a fluency task, which consisted of the number of animals an individual could name in a 1-min interval, forward and backward digit span, and immediate and delayed recall of a 12-item list. The correlations among the individual cognitive components ranged from .33 to .46 and justified the creation of an overall composite (McGue and Christensen 2001). The cognitive composite was defined as the sum of the five standardized components. For all LSADT and MADT assessments, standardization was based on the same set of means and standard deviations (which were based on the 1995 LSADT assessment, which was the first assessment in which the cognitive battery had been administered). In this way the overall cognitive composite is commensurate across assessments. The cognitive composite score has high internal consistency reliability (.75) and is moderately stable over a 2-year interval

(.60). Among the 9,045 participants, we have a valid cognitive composite score for 8,481 (94 %) individuals and 2,238 twin pairs. The major reasons for missing data were proxy response in LSADT and refusal to complete one or more cognitive tests in both surveys.

Grip strength—Hand grip strength in kilograms was measured three times with each hand while the participant was standing, using a Smedley dynamometer (TTMs, Tokyo, Japan). The maximum value across the six readings was taken as the individual's grip strength score. While grip strength was assessed in the initial MADT assessment, it was not included in LSADT until the 1999 assessment. The LSADT grip strength score used in the current study is still based on the initial grip strength assessment of each participant. Because of mortality among LSADT 1995 and 1997 participants, however, the number of participants completing the initial grip strength assessment is somewhat less than those completing the depression or cognitive assessments. Among the 9,045 participants in the two surveys, we have valid grip strength scores on 7,129 (79 %) individuals and 1,855 twin pairs.

Scaling—To facilitate comparison across measures and age groups, scores on each of the three outcome measures were linearly transformed to have a mean of 50 and a standard deviation (SD) of 10 (i.e., a T-score metric) in the youngest age group. Because the transformation was based on the total sample in the youngest age group, transformed scores retained any mean or variance differences due to sex.

Statistical analysis

The effect of age on twin similarity was assessed separately in the MZ and DZ samples using an adaptation of the Defries–Fulker regression model (Defries and Fulker 1985). Specifically, using double entered twin data we fit the following regression equation:

$$T_{ij} = a + b_1 Age_i + b_2 Age_i^2 + b_3 C_{ij} + b_4(Age_i \times C_{ij}) + b_5(Age_i^2 \times C_{ij})$$

where T_{ij} is the observed score for the jth (j = 1,2) twin in the ith (i = 1,… *N*) twin pair, C_{ij} is the cotwin's score, Age_i is the age at the time of assessment for the ith twin pair, and Age_i^2 is the squared pair age. Because members of a twin pair were typically assessed within a few weeks of each other (the difference in twin age averaged less than 1 month for both MZ and DZ twins and in no case exceeded 5 months), we used the average age of the twin pair rather than individual twin age in all analyses. The regression coefficient for the product of cotwin's score with assessment age, b_4, and squared age, b_5, can be used to test the significance of age as a moderator of twin resemblance, as it is relatively easy to show that with double entered data the twin correlation, r_i, is being modeled as a quadratic function of age (Rodgers and McGue 1994):

$$r_i = b_1 + b_4 Age_i + b_5 Age_i^2$$

Standard errors and test statistics reported here have been corrected for the double entry of the twin data.

We also investigated age moderation of the biometric decomposition of variance using Mx (Neale et al. 2004). Specifically, we fit the standard ACE model, where A corresponds to the additive genetic component of variance, C to the shared environmental component, and E to the non-shared environmental component. We allowed each of these components of variance to vary as a linear and quadratic function of twin pair age using the definition variable feature implemented in Mx. Model fit was judged using both a χ^2 goodness-of-fit

test and the Akaike Information Criterion (AIC = $\chi^2 - 2df$), the latter providing a fit statistic that balances model fit with parsimony.

Results

The main effect of age

We first sought to confirm the expectation that each of our three outcomes was significantly associated with age by regressing each of the outcomes on age and squared age. In all cases, the effect of both age and squared age was significant, accounting in total for 12.4, 27.5 and 42.1 % of the variance in the depression symptom, cognitive composite and grip strength scores, respectively. Figure 1 plots the mean outcome score as a function of age for the three outcomes. As is evident, the effect of age is pronounced for all outcomes. From the mid-forties to the mid-eighties (where the vast bulk of the sample lies), depression symptom scores increased by approximately 1.5 SDs, while cognitive and grip strengths scores decreased by about 2 SDs each. While the quadratic effect of age was significant for each of the three outcomes, visual inspection of the plots suggest that the relationship of outcome with age is predominantly linear for each outcome.

Age moderation of twin similarity

Analysis of twin similarity is based on the residuals from regressing each of the three outcomes on age, squared age and sex (McGue and Bouchard 1984). Table 2 gives the overall MZ and DZ twin correlations separately in the male and female samples as well as in the combined male and female sample. As the twin correlations were quite similar in the male and female subsamples, we focus here on the pooled correlations. In every case, the MZ twin correlation is substantially greater than the corresponding DZ twin correlation, implicating the existence of genetic influences. This is expected since we have previously published evidence that each of these three phenotypes is heritable. Next, we plotted the MZ and DZ twin correlations as a function of age group in Fig. 2. To minimize the effect of age groups with small sample size, correlations were computed using a moving window of three consecutive 5-year intervals. While it appears that MZ twin similarity may decline with age for both the cognitive composite and grip strength and increase with age for depression symptomatology, when judged against the plotted standard errors the age differences in the correlations do not appear to be especially pronounced. Regardless, the purpose of the DF and Mx analyses is to provide a formal statistical framework for analyzing the age differences plotted in Fig. 2.

Table 3 gives the squared multiple correlation (R^2) and associated p value from the DF regression analyses for three different regression models. Model 1 included effects for age, squared age and cotwin score. In model 1, cotwin similarity is modeled as a constant over age. Model 2 included all model 1 effects plus the cotwin by age interaction effect. Model 2 thus allowed cotwin similarity to be linearly moderated by age. Finally, model 3 included all the model 2 effects plus the cotwin by squared age interaction effect. Model 3 thus allowed for nonlinear forms of age moderation. For all three outcome phenotypes and for both zygosity groups the pattern of results from the DF analyses is the same. The R^2 from model 1 was significant but the increment in R^2 for both model 2 and model 3 were small and non-significant. The DF analysis thus provides no support for age moderation of twin similarity. Because we failed to find significant age moderation effects on twin similarity, we estimated the power of the DF analyses using the Quanto software program (Gauderman and Morrison 2006). We determined power for the linear age moderation effect, which depends on the size of the sample and the age distribution for each phenotype-zygosity combination. At a two-tailed α of .05, we had power of at least 80 % to detect an increase in the twin correlation of between .021/decade (for depression symptomatology in the DZ twin sample) to .024/

decade (for grip strength in the MZ sample). Over 40 years, which spans the bulk of our sample, we consequently have adequate power to detect linear shifts in twin correlations on the order of .10.

Finally, the results of the Mx analysis are summarized in Table 4. In these analyses, for each outcome the full model included A, C, and E components of variance, each of which was moderated by age and squared age. The first step in our analysis was to compare the ACE Full Moderator model (model 1 in Table 4) with the corresponding CE Full Moderator (model 2) and AE Full Moderator (model 3) models. Results for these model comparisons were very consistent. In every case by both χ^2 test and AIC, the CE model fit the data poorly while the AE model fit the data well. We consequently restricted all subsequent analyses to submodels of the AE model. Within the AE model, we could drop all moderation effects without significant decrements in model fit for both depression symptoms and the cognitive composite. Only for grip strength was there evidence of age moderation, as in this case the AE model with full moderation (model 3) fit better than the AE model without moderation (model 3(a)). Fitting the AE model with a linear age moderator but not a squared age moderator (model 3(b)) produced a model that fit well relative to the full AE model. Consequently, the age moderation effect on the grip strength variance components appears to be primarily linear.

In order to get a sense for how the variance component estimates for the three outcomes differed by age, Fig. 3 gives estimates of the non-standardized A and E components as well as the (standardized) heritability estimate as a function of age. Even though age moderation was statistically significant only for grip strength, estimates given in the figure are based on the AE model with full moderation (i.e., model 3). This allowed us to determine whether there might be a suggestion of age trends even if they were not statistically significant. To facilitate comparison across the variance components, without loss in generality we scaled the total phenotypic variance to be 100 in the youngest age group for each outcome. As is evident, the heritability plots are remarkably flat, suggesting little variability in heritability from age 45 to 85, where the vast bulk of our sample lies. Indeed, heritability estimates across the ages sampled fell within a 5 % interval for depression symptoms, within a 4 % interval for the cognitive composite, and within a 10 % interval for grip strength. Moreover, in no case did the heritability estimates show a monotonic increasing or decreasing pattern with age. Interestingly, the variance components that underlie the estimates of heritability did show an age pattern that varied across the three phenotypes. For the cognitive composite, both the A and E plots are flat over age. However, for depression symptomatology both the A and E components tended to increase with age, while for grip strength, they both tended to decrease. Nonetheless, it is important to emphasize that statistical support for age differences in the variance component estimates was found only with grip strength.

Discussion

Because the previous literature had not found a consistent pattern of age moderation of heritable effects in late-life, we investigated age moderation of twin similarity, and the magnitude of both the raw and standardized biometric variance components in a sample of more than 2,000 pairs of MZ and like-sex DZ twins aged 46–96. For measures of depression symptomatology, overall cognitive ability, and hand grip strength we found little evidence for age-moderated heritable effects. Our findings thus suggest that twin similarity is maintained over a broad range of the lifespan, fully spanning the latter half of life. This is a surprising finding since none of the measures we analyzed was developmentally static; they all evidenced pronounced mean-level changes with age. A priori it seems reasonable to expect that twins would become less similar over time, a consequence of the cumulative

impact of a lifetime of divergent experiences. Indeed, even if only anecdotal, this seems to be the case for physical appearance: based on our experiences, the typical pair of 80-year old MZ twins does not look anywhere as similar as the typical pair of 20-year old MZ twins.

Our findings thus seem inconsistent with earlier research, especially with SATSA, which has reported age differences in the magnitude of both raw and standardized genetic effects (e.g., Finkel et al. 1998; Reynolds et al. 2005). One possible source of the discrepancy between the present report and these earlier reports could be differences between cross-sectional and longitudinal findings. Our analysis is based on cross-sectional data only; these earlier reports were based on a combined analysis of cross-sectional and longitudinal observations. For example, the figures published in the Finkel et al. (1998) study clearly show that the evidence supporting age differences in biometric variance components is much stronger for the longitudinal than for the cross-sectional data. While developmental researchers are generally inclined to favor longitudinal findings over cross-sectional findings, we believe it would be a mistake to do so here especially since our aim is to characterize age *differences* (which cross-sectional studies are designed to investigate) and not age *changes* (which can only be investigated using longitudinal approaches). We focus here on cross-sectional results because we are concerned that practice effects, differential attrition, and small sample size might affect results from longitudinal comparisons. Regardless, the biometric variance components estimated in both cross-sectional and longitudinal twin designs are necessarily based on twin pairs where both participate. Selective attrition of one twin due to bad health or death (both of which increase exponentially with age) could result in only the most similar twin pairs remaining in the study sample, perhaps compensating for growing twin divergence in a way that makes the overall pattern of twin similarity appear to be stable. The degree to which cross-sectional and longitudinal results produce robust differences in large samples is a topic that while beyond the scope of the present investigation is deserving of future investigation.

Alternatively, differences between the present results and those reported by SATSA may reflect differences in the measures used. This may be especially true for the cognitive composite, which in the current study is comprised of memory measures (i.e., word recall and digit span) and a single verbal fluency task, but is based on a broad range of 13 cognitive ability measures in the Swedish study (Finkel et al. 1998). In the Reynolds et al. (2005) study, the heritability estimates for the two memory measures (Thurstone's picture memory and digit span) were generally stable from age 50 to 80, which is consistent with our failure to find differences in heritability for our cognitive composite measure. Nonetheless, the variance components that underlie the heritability estimates were not stable across age in the Swedish study, which is not consistent with our results. Specifically, in the Swedish study the estimate of the genetic component of variance increased with age for both memory measures, while the estimate of the non-shared environmental component of variance increased markedly with age for the picture memory task but decreased with age for digit span. We can only speculate as to whether a cognitive composite score based on a diverse set of cognitive scores, such as those used in SATSA, might have evidenced a different pattern of age effects than that found with the cognitive composite used in the current study.

If our findings prove to be an accurate representation of adult development, it leaves us with the question as to how twins are able to maintain their psychological similarity over periods of 30 years or more. One possibility is that new genetic variation, perhaps the genetic variation predicted to exist by the mutation accumulation model, almost perfectly compensates for the loss of similarity due to the cumulative impact of differences in the twins' experiences. While we believe this remains a viable hypothesis, only one of our phenotypes showed a pattern of findings consistent with this hypothesis. For depression

symptomatology, age-related increases in non-shared environmental effects were nearly perfectly matched by age-related increases in genetic variance to produce stable heritability estimates. Moreover, this explanation would lead to the prediction that genetic factors contribute to behavioral change in adulthood. There is, however, only limited evidence of genetic contributions to behavioral changes in late-life. For example, in the cognitive domain we estimated the heritability of the slope parameter (a measure of linear change over the multiple waves of assessment) for the general cognitive composite in LSADT to be only 6 % (McGue and Christensen 2002). In a similar analysis of SATSA longitudinal data, Reynolds et al. (2005) reported the heritability of the slope to be 1 % for their first principal component cognitive measure. However, these investigators reported a heritability estimate of 43 % for the quadratic component of change, indicating that non-linear patterns of change may be heritable. Finally, in a large sample of more than 1,200 male veteran twin pairs, Lyons et al. (2009) reported an estimate of 17 % for the heritability of change over an approximately 35-year interval (age 20 to the mid-50s). The one notable exception to the general finding of modest genetic contributions to behavioral change in adulthood comes from the recent study by McArdle and Plassman (2009), who reported strong genetic contributions to change in memory performance over 12 years in the NSA–NRC Twin Registry of male WWII veterans. In any case, there is a clear need for additional longitudinal investigations of the contribution of genetic factors to behavioral change, especially long-term longitudinal studies that span decades rather than just years of life.

A second explanation for the maintenance of twin similarity over adulthood, the explanation we favor, is that it is fundamentally environmental in origin. There is now a substantial behavioral genetic literature documenting heritable contributions to environmental measures (Plomin and Bergeman 1991). Using LSADT data, we have previously shown that twins, and especially MZ twins, tend to create similar social (McGue and Christensen 2007), and physical (Frederiksen and Christensen 2003) environments. Twins may be able to maintain their behavioral similarity in part because they are able to construct environments that complement and reinforce those heritable characteristics (c.f., Scarr and McCartney 1983). Alternatively, we might expect declines in twin similarity when the infirmaries of late-life prevent twins from constructing similar environments. From this perspective, the onset of late-life disabilities rather than the attainment of more years of age may be key to understanding changes in twin similarity, a possibility that has not been addressed in the existing literature.

As with any single study, ours is not without its limitations, which should be taken into account when interpreting our results. First, while our sample spans ages 46–96, the number of twin pairs older than age 85 (total number of MZ and DZ pairs <40 for all three phenotypes) is limited. We are not in position to draw conclusions about twin similarity in very late-life. Other research has shown that the heritability of lifespan in very late-life differs, albeit being greater, than the heritability of lifespan at earlier life stages (Hjelmborg et al. 2006). Nonetheless, our findings cannot rule out the possibility that for the outcomes we consider twin similarity declines in very late-life. Second, even though our sample is large, we cannot rule out small differences in heritability or twin similarity across age. For example, estimates of the heritability of grip strength varied by as much as 9 % over the age range we sampled. This difference is not large, nor was it directional as the minimal heritability estimate occurred in the late-60s and was higher at either younger or older ages. Third, our analysis fails to take into account the shifting pattern of mortality with age. It may be that the apparent conservation of twin similarity across age is a consequence of greater mortality among dissimilar as compared to similar twin pairs. Finally, as previously noted we have investigated only three phenotypes. Although depression symptoms, cognitive functioning, and grip strength represent three key aging outcomes, we recognize that age

patterns of twin similarity and heritability may be different for other phenotypes. We hope our paper encourages others to explore this intriguing issue.

In summary, analysis of twin similarity for a composite measure of cognitive functioning, depression symptomatology, and grip strength in a large sample of twin pairs aged 46–96 failed to provide consistent evidence of age declines in twin similarity. How twins maintain their similarity over many years remains to be determined but may be a consequence of novel genetic variance, which counterbalances the cumulative impact of environmental effects, or the ability of twins to create environments that reinforces and maintains their behavioral similarity. Regardless, we hope our findings encourage others to systematically explore the nature of twin resemblance over the latter half of the lifespan in much the same way that it has been explored at earlier developmental stages.

Acknowledgments

References

Bergen SE, Gardner CO, Kendler KS. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. Twin Res Hum Genet. 2007; 10(3):423–433. [PubMed: 17564500]

Carmelli D, Swan GE, Kelly-Hayes M, Wolf PA, Reed T, Miller B. Longitudinal changes in the contribution of genetic and environmental influences to symptoms of depression in older male twins. Psychol Aging. 2000; 15(3):505–510. [PubMed: 11014713]

Christensen K, Holm NV, McGue M, Corder L, Vaupel JW. A Danish population-based twin study on general health in the elderly. J Aging Health. 1999; 11:49–64. [PubMed: 10848141]

Defries JC, Fulker DW. Multiple regression analysis of twin data. [Article]. Behav Genet. 1985; 15(5): 467–473. [PubMed: 4074272]

Eaves L, Martin N, Heath A, Schieken R, Meyer J, Silberg J, Corey L. Age changes in the causes of individual differences in conservatism. Behav Genet. 1997; 27(2):121–124. [PubMed: 9145550]

Finkel, D.; Reynolds, CA. Behavioral genetic investigations of cognitive aging. In: Kim, YK., editor. Handbook of behavior genetics. Springer; New York: 2010. p. 101-112.

Finkel D, Pedersen NL, McGue M, McClearn GE. Heritability of cognitive abilities in adult twins: comparison of Minnesota and Swedish data. Behav Genet. 1995; 25:421–431. [PubMed: 7487839]

Finkel D, Pedersen NL, Plomin R, McClearn GE. Longitudinal and cross-sectional twin data on cognitive abilities in adulthood: the Swedish adoption/twin study of aging. Dev Psychol. 1998; 34:1400–1413. [PubMed: 9823519]

Finkel D, Pedersen N, Reynolds CA, Berg S, de Faire U, Svartengren M. Genetic and environmental influences on decline in biobehavioral markers of aging. Behav Genet. 2003; 33:107–123. [PubMed: 14574146]

Frederiksen H, Christensen K. The influence of genetic factors on physical functioning and exercise in second half of life. Scand J Med Sci Sports. 2003; 13:9–18. [PubMed: 12535312]

Frederiksen H, Gaist D, Petersen HC, Hjelmborg J, McGue M, Vaupel JW, Christensen K. Hand grip strength: a phenotype suitable for identifying genetic variants affecting mid- and late-life physical functioning. Genet Epidemiol. 2002; 23:110–122. [PubMed: 12214305]

Gaist D, Bathum L, Skytthe A, Jensen TK, McGue M, Vaupel JW, Christensen K. Strength and anthropometric measures in identical and fraternal twins: no evidence of masculinization of females with male co-twins. Epidemiology. 2000; 11(3):340–343. [PubMed: 10784255]

Galton F. The history of twins, as a criterion of the relative powers of nature and nurture. Fraser's Mag. 1875; 12:566–576.

Gatz M, Pedersen NL, Plomin R, Nesselroade JR, McClearn GE. Importance of shared genes and shared environments for symptoms of depression in older adults. J Abnorm Psychol. 1992; 101:701–708. [PubMed: 1430610]

Gauderman, WJ.; Morrison, JM. [Accessed 12 May 2012] QUANTO 1.1: a computer program for power and sample size calculations for genetic epidemiology studies. 2006. http://hydra.usc.gxe

Hamilton WD. The moulding of senescence by natural selection. J Theor Biol. 1966; 12:12–45. [PubMed: 6015424]

Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, Christensen K. Genetic influence on human lifespan and longevity. Hum Genet. 2006; 119(3):312–321. [PubMed: 16463022]

Johnson W, McGue M, Gaist D, Vaupel JW, Christensen K. Frequency and heritability of depression symptomatology in the second half of life: evidence from Danish twins over 45. Psychol Med. 2002; 32(7):1175–1185. [PubMed: 12420887]

Kjøller, M. Health and morbidity in Denmark, 1994. Danish Institute for Clinical Epidemiology; Copenhagen: 1996.

Koenig LB, McGue M, Iacono WG. Stability and change in religiousness during emerging adulthood. Dev Psychol. 2008; 44(2):532–543. [PubMed: 18331142]

Lyons MJ, York TP, Franz CE, Grant MD, Eaves LJ, Jacobson KC, Kremen WS. Genes determine stability and the environment determines change in cognitive ability during 35 years of adulthood. Psychol Sci. 2009; 20(9):1146–1152. [PubMed: 19686293]

McArdle JJ, Plassman BL. A biometric latent curve analysis of memory decline in older men of the NSA–NRC Twin Registry. Behav Genet. 2009; 39(5):472–495. [PubMed: 19404731]

McCartney K, Harris MJ, Bernieri F. Growing up and growing apart: a developmental meta-analysis of twin studies. Psychol Bull. 1990; 107(2):226–237. [PubMed: 2138795]

McGue M, Bouchard TJ. Adjustment of twin data for the effects of age and sex. Behav Genet. 1984; 14:325–343. [PubMed: 6542356]

McGue M, Christensen K. Genetic and environmental contributions to depression symptomatology: evidence from Danish twins 75 years of age and older. J Abnorm Psychol. 1997; 106:439–448. [PubMed: 9241945]

McGue M, Christensen K. The heritability of cognitive functioning in very old adults: evidence from Danish twins aged 75 years and older. Psychol Aging. 2001; 16:272–280. [PubMed: 11405315]

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; 28:435–452. [PubMed: 12227922]

McGue M, Christensen K. Social activity and healthy aging: a study of aging Danish twins. Twin Res Hum Genet. 2007; 10(2):255–265. [PubMed: 17564515]

McGue, M.; Bouchard, TJ.; Iacono, WG.; Lykken, DT. Behavioral genetics of cognitive ability: a life span perspective. In: Plomin, R.; McClearn, GE., editors. Nature, nurture and psychology. Americal Psychological Association; Washington DC: 1993. p. 59-76.

Neale, MC.; Boker, SM.; Xie, G.; Maes, HH. Mx: statistical modeling. 6. Department of psychiatry; Box 126 MCV, Richmond: 2004.

Neiss M, Almeida DM. Age differences in the heritability of mean and intraindividual variation of psychological distress. Gerontology. 2004; 50(1):22–27. [PubMed: 14654723]

Platz, M. The elderly in their homes: Living conditions. Vol. 1. The Danish National Institute of Social Research; Copenhagen: 1989.

Platz, M. The elderly in their homes: How do they cope?. Vol. 2. The Danish National Institute of Social Research; Copenhagen: 1990.

Plomin R, Bergeman CS. The nature of nurture: genetic influence on environmental measures. Behav Brain Sci. 1991; 14(3):373–385.

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. Dev Psychol. 2005; 41(1):3–16. [PubMed: 15656733]

Rodgers JL, McGue M. A simple algebraic demonstration of the validity of DeFries-Fulker analysis in unselected samples with multiple kinship levels. Behav Genet. 1994; 24(3):259–262. [PubMed: 7945155]

Roth M, Tym E, Mountjoy CQ, Huppert FA, Hendrie FA, Verma S, Goodard R. CAMDEX: a standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. Br J Psychiatry. 1986; 149:698–709. [PubMed: 3790869]

Scarr S, McCartney K. How people make their own environments: a theory of genotype-environment effects. Child Dev. 1983; 54(2):424–435. [PubMed: 6683622]

Skytthe A, Kyvik K, Holm NV, Vaupel JW, Christensen K. The Danish twin registry. Twin Res. 2002; 5:352–357. [PubMed: 12537858]

Whitfield KE, Edwards CL, Brandon D, McDougald C. Genetic and environmental influences on depressive symptoms by age and gender in African American twins. Aging Ment Health. 2008; 12(2):221–227. [PubMed: 18389402]
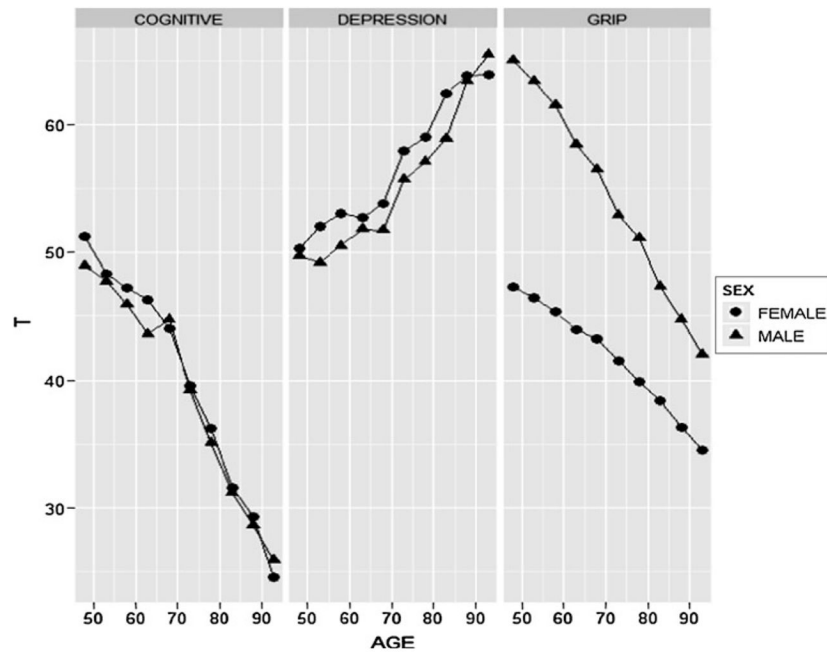
Fig. 1.
Mean outcome score as a function of age. Plotted are mean scores for 5-year intervals ranging from 45–50 to 91–96. Scores are scaled to have a mean of 50 and SD of 10 in the youngest age group (i.e., a T-score metric). In all sex-age group combinations except males in the oldest age group, standard error of the mean is less than 1 point. For males in the oldest age group, standard error of the mean was between 1 and 2
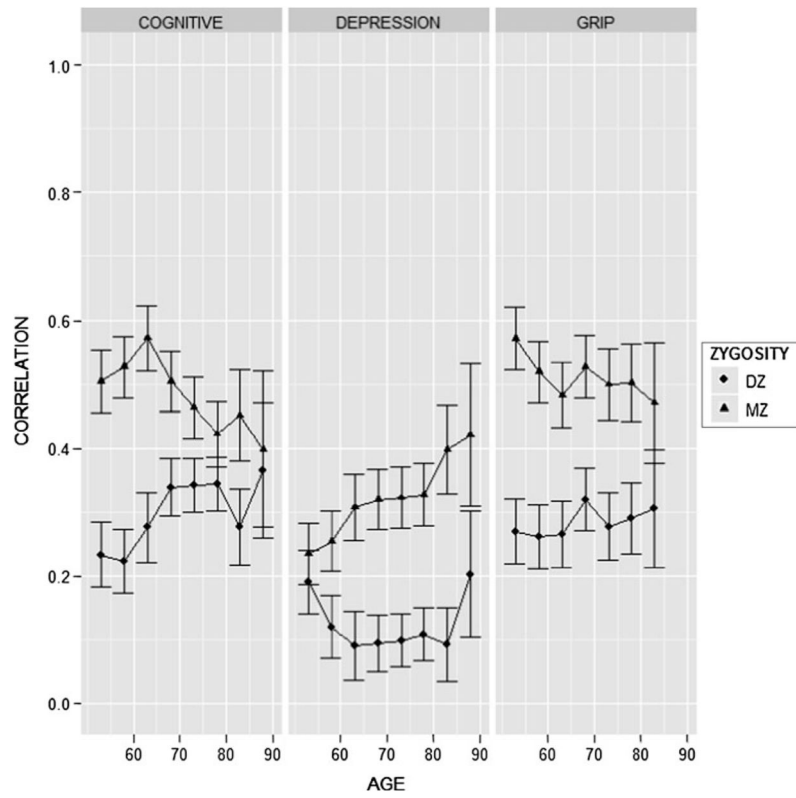
Fig. 2.
MZ and DZ twin correlations for outcome as a function of age. Correlations were computed using a moving window of three consecutive 5-year intervals and plotted for the midpoint of that interval. *Error bars* give ± one standard error of the correlation estimate
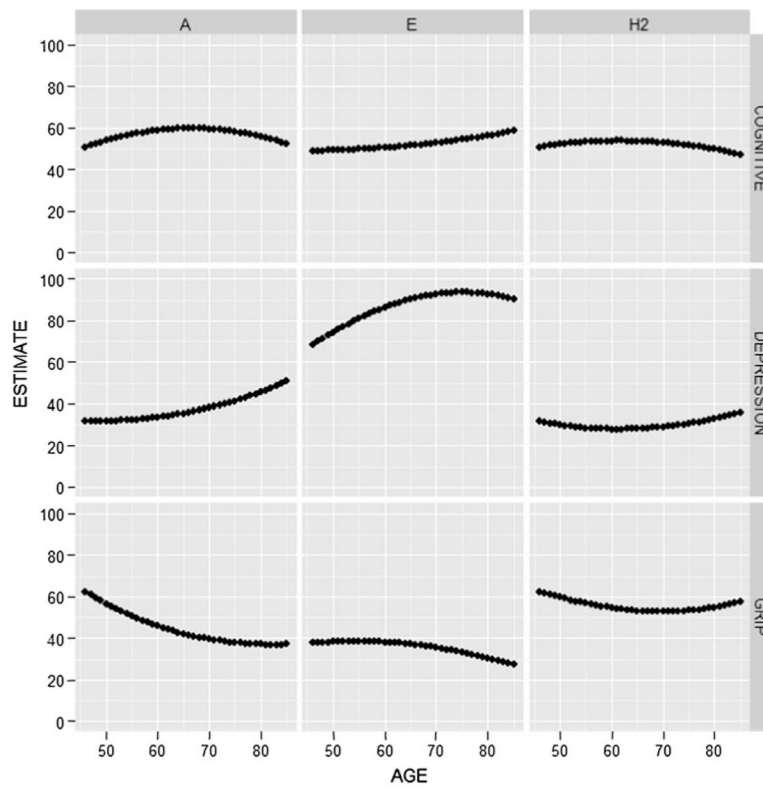
Fig. 3.
Estimates of additive genetic (A), non-shared environmental (E), and heritability (H2) as a function of age for the three outcomes. To facilitate comparisons across the estimates, total phenotypic variance was scaled to equal 100 in the youngest age for each outcome. Estimates are based on the AE model with full age moderation (model 3 in Table 4). Only for grip strength is there evidence of significant age differences in the variance component estimates

Table 1

Number of participating twin pairs by age group

	Male		Female		Total
	MZ	**DZ**	**MZ**	**DZ**	
46–50	61	62	63	58	244
51–60	153	144	147	133	577
61–70	121	106	119	100	446
71–80	152	214	213	331	910
81–90	19	19	50	61	149
>90	2	0	2	2	6
Total	508	545	594	685	2,332

Table 2

Twin correlations in the overall sample

	Male		Female		Total	
	MZ	**DZ**	**MZ**	**DZ**	**MZ**	**DZ**
Depression symptoms	.29 (N = 502)	.14 (N = 537)	.32 (N = 590)	.11 (N = 678)	.31 (N = 1,092)	.12 (N = 1,215)
Cognitive composite	.52 (N = 486)	.29 (N = 522)	.47 (N = 575)	.29 (N = 645)	.50 (N = 1,061)	.29 (N = 1,167)
Grip strength	.54 (N = 452)	.30 (N = 454)	.55 (N = 484)	.29 (N = 465)	.55 (N = 936)	.30 (N = 919)

All twin correlations have been adjusted for the main effects of age, squared age and sex

Table 3

Squared multiple correlations for DeFries–Fulker moderated regression analyses

		Model 1		Model 2		Model 3	
		R^2	p	ΔR^2	p	ΔR^2	p
Depression symptoms	MZ	.187	<.001	.001	>.05	.000	>.05
	DZ	.105	<.001	.001	>.05	.000	>.05
Cognitive composite	MZ	.406	<.001	.001	>.05	.000	>.05
	DZ	.275	<.001	.002	>.05	.000	>.05
Grip strength	MZ	.508	<.001	.000	>.05	.000	>.05
	DZ	.382	<.001	.000	>.05	.000	>.05

Model 1 included age, squared age, and cotwin effect, model 2 included model 1 terms plus the interaction of the cotwin effect with age, and model 3 included model 2 terms plus the squared age effect

Table 4

Fit statistics for the age-moderated models

Model	Depression symptoms					Cognitive composite					Grip strength				
	−2lnL	df	χ^2 (df)	p	AIC	−2lnL	df	χ^2 (df)	p	AIC	−2lnL	df	χ^2 (df)	p	AIC
1. ACE: full age moderation[a]	34,315.9	4,627				33,433.0	4,541				27,055.4	3,708			
2. CE: full age moderation	34,338.8	4,630	22.9 (3)	<.001	16.9	33,501.6	4,544	58.6 (3)	<.001	52.6	27,113.1	3,711	57.7 (3)	<.001	51.7
3. AE: full age moderation	34,317.0	4,630	1.1 (3)	.78	−4.9	33,450.7	4,544	7.7 (3)	.05	1.7	27,059.2	3,711	3.8 (3)	.28	−2.2
3(a). AE: no age moderation	34,338.9	4,634	2.3 (7)	.94	−4.7	33,454.2	4,548	11.2 (7)	.13	−2.8	27,083.0	3,715	27.6 (7)	<.001	13.6
3(b). AE: no squared age moderation	34,318.5	4,632	2.6 (5)	.76	−7.4	33,451.4	4,546	8.4 (5)	.14	−1.6	27,060.6	3,713	5.2 (5)	.39	−4.8
3(c). AE: no age moderation on A	34,319.4	4,632	3.5 (5)	.62	−6.5	33,451.4	4,546	8.4 (5)	.14	−1.6	27,069.3	3,713	13.9 (5)	.02	3.9
3(d). AE: no age squared moderation on A	34,317.2	4,631	1.3 (4)	.86	−6.7	33,451.3	4,545	8.3 (4)	.08	0.3	27,060.2	3,712	4.8 (4)	.31	−3.2
3(e). AE: no age moderation on E	34,322.9	4,632	7.0 (5)	.22	−3.0	33,452.1	4,546	9.1 (5)	.11	−0.9	27,063.2	3,713	7.8 (5)	.17	−2.2
3(f). AE: no age squared moderation on E	34,318.3	4,631	2.4 (5)	.66	−5.6	33,450.8	4,545	7.8 (4)	.10	−0.2	27,060.2	3,712	4.8 (4)	.31	−3.2

Models 3(a) through 3(f) are submodels of model 3

[a] Base model for computing χ^2 difference statistic and AIC