Psychology and Aging
2006, Vol. 21, No. 1, 62–73

0882-7974/06/$12.00 DOI: 10.1037/0882-7974.21.1.62

Age and Sex Differences in Reaction Time in Adulthood: Results From the United Kingdom Health and Lifestyle Survey

Geoff Der
University of Glasgow

Ian J. Deary
University of Edinburgh

Reaction times (RTs) slow and become more variable with age. Research samples are typically small, biased, and of restricted age range. Consequently, little is known about the precise pattern of change, whereas evidence for sex differences is equivocal. The authors reanalyzed data for 7,130 adult participants in the United Kingdom Health and Lifestyle Survey, originally reported by F. A. Huppert (1987). The authors modeled the age differences in simple and 4-choice reaction time means and variabilities and tested for sex differences. Simple RT shows little slowing until around 50, whereas choice RT slows throughout the adult age range. The aging of choice RT variability is a function of its mean and the error rate. There are significant sex differences, most notably for choice RT variability.

Keywords: reaction time, cognitive aging, sex differences, adulthood, Health and Lifestyle Survey (HALS)

The use of reaction times (RTs) as measures of cognitive functioning has a long history, dating back at least to the 19th century when Galton used them as part of the battery of tests included in his "anthropometric laboratory" (Pearson, 1924). Over a century later, their appeal remains undiminished, particularly in research on aging. There are a number of reasons for this. One is their relative simplicity. RT tasks are simpler to devise and administer than most other cognitive measures or psychometric tests. Nonetheless, they are commonly found to be correlated with other cognitive measures and sometimes to be better predictors of important outcomes. For example, in a recent study (Deary & Der, 2005a), we found RTs to be better predictors of mortality than scores on the Alice Heim 4 Test of General Intelligence (Heim, 1970).

Empirically, RTs are strongly associated with age. It is well established that, during adulthood, RTs increase and become more variable with age. Galton's own data provided some of the earliest evidence for slowing RTs (Johnson et al., 1985; Koga & Morant, 1923). Fozard, Vercruyssen, Reynolds, Hancock, and Quilter (1994) and Deary and Der (2005b) summarized the support from cross-sectional and longitudinal data, respectively. This decline parallels age-related declines in other areas of cognitive functioning. In a meta-analysis of the relationship of age to a range of cognitive measures, Verhaeghen and Salthouse (1997) found a weighted average correlation between age and RT of .52; they also found that "between 71% and 79% of the age-related variance in the cognitive variables was shared with speed" (p. 246), which included RT. From these and other results, they concluded that "the speed variable may deserve special status in the context of cognitive aging" (p. 246).

Madden (2001) described processing speed as "a fundamental property of the central nervous system" (p. 288), and RTs are measures of this processing speed. This viewpoint suggests a simple and parsimonious explanation for age-related cognitive decline, namely, that it is due to a general slowing of the system's processing speed (Salthouse, 1996). However, this processing speed hypothesis of cognitive aging is not without its critics. Sliwinski and Buschke (1999) showed that controlling for processing speed reduces cross-sectional age effects much more that it does longitudinal effects. They concluded that processing speed is likely to reflect stable, non-age-related, individual differences.

Thus, despite the long research tradition and the wealth of evidence linking RT and age, important aspects of the relationship remain unclear. It has not been definitively established whether the relationship is linear throughout the adult age range. If it is nonlinear, is there evidence for a threshold at which cognitive aging begins? Indeed, it is not even clear whether there is a single pattern for the relationship of RT and age. If a slowing of general processing speed is at the root of cognitive decline, then RT measures would be expected to show a broadly similar age pattern, even if other cognitive measures, such as measures of crystallized intelligence, do not. Most of the evidence concerns mean RTs, and much less is known about age differences in RT variability. Finally, it is not clear whether there are consistent sex differences in RT, and if so, what part they would play in theories of cognitive aging.

There are several reasons for these gaps in current knowledge. Many of the samples studied consist entirely of older people. A large part of the evidence for age changes is derived from studies in which an older group was compared with a younger one, rather

Geoff Der, Medical Research Council Social and Public Health Sciences Unit, University of Glasgow, Scotland; Ian J. Deary, Department of Psychology, University of Edinburgh, Scotland.

Supplementary material is available online at http://dx.doi.org./10.1037/0882-7974.21.1.62).

Geoff Der is funded by the Medical Research Council. Ian Deary is the recipient of a Royal Society-Wolfson Research Merit Award.

Correspondence concerning this article should be addressed to Geoff Der, Medical Research Council Social and Public Health Sciences Unit, University of Glasgow, 4 Lilybank Gardens, Glasgow, United Kingdom G12 8RZ. E-mail: geoff@msoc.mrc.gla.ac.uk

than from data covering the whole age range. Hofer and Sliwinski (2001) criticized this type of design as being prone to population level confounds. One such source of confounding arises when the older and younger groups are sampled from different populations, for example, when the younger group comprises students (Botwinick & Storandt, 1974; Harkins, Nowlin, Ramm, & Schroeder, 1974; Mathey, 1976).

Samples studied are frequently small, and hence lack power. The meta-analysis of Verhaeghen and Salthouse (1997), although restricted to studies reporting an age–speed correlation, illustrated the range of typical sample sizes. Verhaeghen and Salthouse amassed a total sample size of just over 11,000 from 28 studies. The largest of these samples, that of Crook and West's (1990) study, had a sample size of 1,205. The samples of only three other studies exceeded 400. Crook and West's study was principally concerned with age-related declines in name recall. RTs were only used as predictors.

Of the large samples used to study RTs, few are representative of the general population. Galton's own subjects were a highly selected group: They had to pay to be tested and as a result "a sizable portion. . .consisted of professionals, semiprofessionals and students" (Johnson et al., 1985, p. 876). The subjects in the Baltimore Longitudinal Study of Aging (BLSA) were from the "upper-middle socioeconomic level," and "about 73% of the men and 64% of the women had at least one college degree" (Fozard et al., 1994, p. 180).

Attempts at combining data from different studies are undermined by the wide variety of different procedures and RT devices used, some of them idiosyncratic. For example, the BLSA used a nonstandard auditory task not originally designed for the collection of RTs.

What is clearly needed is evidence from a large, representative sample covering the whole adult age range and assessed using a standard RT device. The aim of this study was to examine some of the issues highlighted by reanalyzing data from just such a study, the United Kingdom Health and Lifestyle Survey (HALS), which we believe to be the largest population-based sample, covering the whole adult age range, for which RT data are available. More specifically, we aimed to delineate precisely and compare the normal aging of four measures of RT: the mean and variability of both simple and four-choice RTs. At the same time, we modeled and tested for sex differences.

We also conducted additional analyses of RT variability controlling for the mean. In comparison to mean RTs, the variability of RT has received little attention until recent times. Anstey (1999) found that RT mean and standard deviation were similarly related to age, lung function, vibration sense, and grip strength and argued that RT variability is worthy of more attention in cognitive aging research. Hultsch and MacDonald (2004) reviewed the evidence linking cognitive aging to three types of RT variability: interindividual variability, intraindividual variability across tasks, and intraindividual variability within tasks, which they referred to as "diversity," "dispersion," and "inconsistency," respectively. Inconsistency was our main focus, and inconsistency is known to be elevated in a number of neurological conditions, including, for example, mild dementia (Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000) and chronic fatigue syndrome (Fuentes, Hunter, Strauss, & Hultsch, 2001).

Rabbitt, Osman, Moore, and Stollery (2001) suggested another reason for studying RT variability. They claimed that people's fastest choice RTs are relatively unaffected by age but that "substantial individual differences in mean CRTs [choice RTs] often reflect differences in the numbers of unnecessarily slow responses that people make" (p. 982). The implication is that the increase in mean RT with age is a result of increasing variability, and variability is, therefore, a more important component of cognitive aging. They also found inconsistency to be associated with scores on the Culture Fair Intelligence Test (Cattell, 1950). Hultsch, MacDonald, and Dixon (2002) reported a similar finding: There were no age-related increases in inconsistency for the fastest 20% of response latencies, but marked increases for the slowest 20% of responses.

The data for our analysis are drawn from the HALS. According to the introduction, the main HALS report aimed to be "accessible to as wide a readership as possible" and therefore presented "a descriptive account of the information. . .principally in tabular form" and "without the use of statistics" (Cox et al., 1987). In keeping with this brief, the presentation of the RT data included summary statistics (means and standard errors) tabulated by sex and 10-year age bands. On the basis of the tabulated data, Huppert (1987) concluded that age was associated with longer RTs and more RT variability, both between and within subjects. She also noted different patterns of aging for simple and choice RT and "consistent but very small sex differences" (p. 44).

The data from HALS are available for secondary analysis from the United Kingdom Economic and Social Research Council Data Archive. We reanalyzed the data, modeling RTs as a continuous function of age and explicitly testing for sex differences. Because RT mean and variability are correlated, those analyzing the age patterning of RT variability typically control for the mean in one of two ways: either by examining the coefficient of variation (RT *SD*/RT *M*) or by analyzing RT variability controlling for the mean. The coefficient of variation provides a simple summary measure that is useful for comparing the variability in different distributions. Because one of our aims was to compare the age patterns of simple and four-choice RTs, it was an obvious candidate. In contrast, controlling for the mean, although more complex, allows greater flexibility in determining the relationship between the variability and mean and thus more precise control for its influence. This is clear when one considers that modeling the coefficient of variation is equivalent to incorporating the reciprocal of the RT mean as an offset in a model of the RT variability. It is assumed to be the correct functional form without the assumption being tested. Below, we describe both approaches and compare the results.

Method

Participants

The HALS was set up with the aim of exploring the relationship of lifestyles, behaviors, and circumstances to physical and mental health of a large representative sample of the adult British population. Household addresses were randomly selected from electoral registers using a three-stage clustered design. From each household selected, one adult was chosen using a standard sampling technique. The resulting sample comprised 9,003 British adults interviewed between autumn 1984 and summer 1985. The interviews were conducted in the respondents' homes over the course of two visits. An RT task was included in the second visit in which 7,414 respondents took part. These two samples were compared with the national census on a range of characteristics. The single and divorced/separated were found to be slightly underrepresented, and those with the least education and lowest incomes were less likely to take part in the second visit. Nonetheless Blaxter (1987) concluded that "these sources of bias are small and the study appears to offer a good and representative sample of the population" (p. 1).

Measures

RT was measured using a portable device designed especially for the study. A diagram and full description are given in Deary, Der, and Ford (2001). Briefly, the device has an LCD display screen beneath which are five keys labeled *1*, *2*, *0*, *3*, and *4*. The central *0* key is used for the simple RT task. The index finger of the respondent's preferred hand rests on this key, and the respondent is told to press it as quickly as possible after 0 appears in the display. Eight practice trials were followed by 20 test trials. The keys labeled *1*, *2*, *3*, and *4* are used for the four-choice RT task. The respondent rests the index and middle finger of each hand on the keys and presses the corresponding key when one of the four digits appears in the display. There were 8 practice trials and 40 test trials. In the test trials, the digits 1 through 4 each appeared 10 times in a randomized order. For both tasks, the time between the response and the display of the next digit varied randomly between 1 s and 3 s.

The device does not store the results of individual trials but calculates the mean and standard deviation of the test trials in milliseconds. For the choice RT task, the mean and standard deviation of the correct and incorrect responses were recorded separately, and the number of errors was also recorded. The primary focus of this study was on four measures: the mean and standard deviation of the simple RTs and those of the correct responses to the choice RT task.

Analysis

The analysis had two parts. First, we modeled the change in the RT measures over the full age range, while also allowing for any gender differences. Then, we examined in more detail the intraindividual variability. All the analyses were performed using SAS software (Version 8.2, SAS Institute, Cary, NC). For the principal analyses, the general linear models procedure (PROC GLM) is used to fit polynomial regression models. For these models, age was centered on the mean (44.9 years), and the polynomial terms derived from the centered value. We also used the Box-Cox transformation (Box & Cox, 1964) to normalize and stabilize variance (PROC TRANSREG) and nonparametric regression (Hastie & Tibshirani, 1990) with cubic spline smoothers to check the functional form of some of the relationships that we found (PROC GAM).

Results

Of the 7,414 respondents who took part in the second interview, 7,216 completed the RT task. The reasons that the task was not completed were problems with the use of hands (25), poor eyesight (20), equipment failure (78), and miscellaneous other reasons (75). In addition, we excluded 81 cases in which eight or more errors were made on the choice RT task and 5 cases in which the recorded standard deviation in one of the tasks was less than 10 ms. An error rate of 20% suggests problems in correctly carrying out the task and is the same cutoff that was used in another study using the same RT device (Deary & Der, 2005b). This left a working sample of 7,130 respondents with an age range between 18 and 94 years. Although the whole working sample was included in all statistical models, we excluded those who were 82 years or older from graphs of the results because the small numbers and consequent variability tended to unduly influence the scale of the graphs.

Figure 1 shows the means and standard errors of the four RT measures by sex and age in 2-year bands. The corresponding means and standard deviations are given in Table S1 which is available at http://dx.doi.org./10.1037/0882-7974.21.1.62. The male and female results have been offset slightly on the *x*-axis so that the overlap of the standard error bars can be seen more clearly. As expected, both simple and choice RTs slow as people age and become more variable. However, the patterns appear to be different. Simple RT shows little slowing until the 40s, and there is even some suggestion that the intrasubject standard deviation (ISD), which is a measure of intrasubject variability, decreases until the mid 30s. Choice RT, in contrast, shows slowing throughout the age range with corresponding increases in ISD. There is some evidence of gender differences for each measure, with RTs in women being slower and more variable. The largest and most consistent difference across the range of ages is the increased choice RT ISD of women.

Preliminary models were fitted to the data and the residuals examined. The results of the preliminary models and more details of their residuals are given in Table S2 and Figure S1 in the online supplement (available at http://dx.doi.org./10.1037/0882-7974.21.1.62). These revealed positive skewness throughout the age range, decreasing slightly at older ages, together with increased variances at older ages. The results would, therefore, have been biased and difficult to interpret. One way to reduce the biases caused by skewness or nonconstant variance is to transform the data. In practice, the choice of transformation is often made by trying some of the common forms (e.g., the log, square root, or reciprocal) and choosing the one with the best results. Box and Cox (1964) suggested a more rigorous procedure. They considered the family of transformations that can be defined as $y^{(\lambda)} = (y^{\lambda} - 1)/\lambda$ for $\lambda \neq 0$ and $\log(y)$ for $\lambda = 0$ and suggested fitting a range of values for λ and choosing the one that yields the maximum likelihood for the model in question, in this case a polynomial regression of RT on age and sex. The resulting transformation is optimal in two senses: It is chosen from a continuous range of parameter values, as opposed to a few discrete values, and it is optimized for the model to be fitted. The values of λ obtained and used to transform the data were RT mean, −1.65; RT *SD*, −0.36; CRT mean, −1.25; and CRT *SD*, −0.31.

The transformed data were then modeled as before. The summary statistics for these models of the transformed data are given in Table 1 of this article and in Figure S2 in the online supplement (available at http://dx.doi.org./10.1037/0882-7974.21.1.62). Figure 2 shows the predicted values and 95% confidence intervals for these models transformed back to the original units. To aid comparison between this figure and Figure 1, we plotted each RT measure on the same scale in both. The most notable difference is that the curves for both simple RT measures are shallower in Figure 2 with predicted means for the oldest subjects much lower than the observed means in Figure 1. In comparison, the curves for the choice RT measures are not markedly different from those in Figure 1. The gender differences already noted remain, with the biggest difference for choice RT variability and the smallest for choice RT mean. Women's mean RTs, both simple and choice, appear to slow more rapidly at older ages.

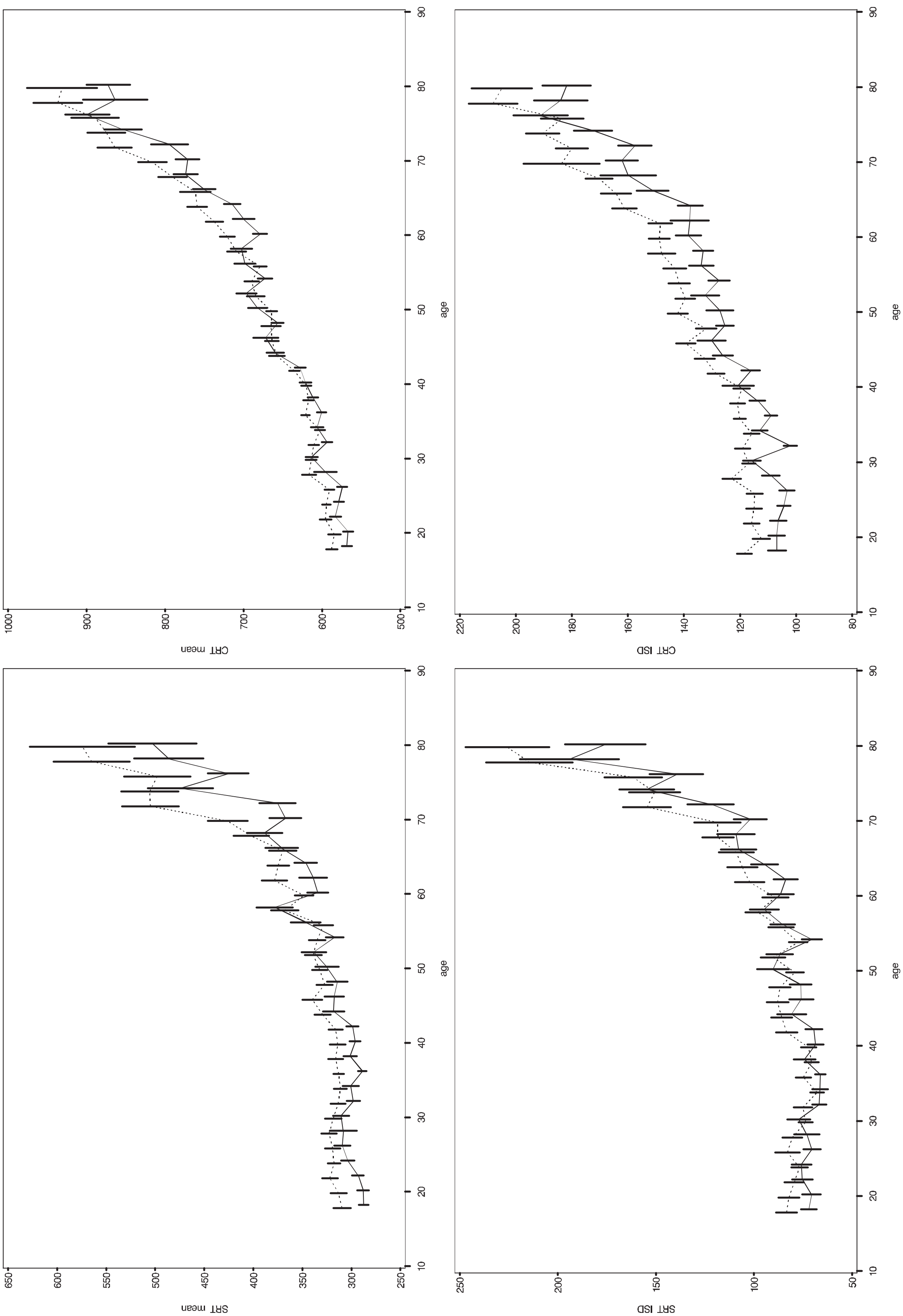

Figure 1. Means and standard errors of reaction time measures (in milliseconds) by age and sex. Solid line = men; dashed line = women. The results for men and women have been offset slightly on the *x*-axis so that the overlap of the standard error bars can be seen more clearly. SRT = simple reaction time; CRT =choice reaction time; ISD = intrasubject standard deviation.

Table 1
Summary Statistics for Models of Box–Cox Transformed Reaction Time (RT) Measures

Outcome	Parameter	Estimate	*SE*	*F*	$p > F$	η^2
SRT mean	Intercept	0.6060111935	3.6921508E-7			
	Age	0.0000002515	2.523713E-8	1,007.76	<.0001	
	Age^2	0.0000000097	1.0046087E-9	205.10	<.0001	.145
	Age^3	0.0000000001	3.556609E-11	9.91	.0017	
	Sex	−0.0000029779	5.4592303E-7	100.64	<.0001	.012
	Age*sex	0.0000000044	2.3858424E-8	0.34	.5603	
	Age^2*sex	−0.0000000031	1.3219769E-9	5.56	.0184	.001
SRT ISD	Intercept	2.148890098	0.00238173			
	Age	0.001594583	0.00008856	594.16	<.0001	
	Age^2	0.000079802	0.00000488	264.06	<.0001	.107
	Sex	−0.016784090	0.00288069	33.95	<.0001	.004
CRT mean	Intercept	0.7997475115	9.6802073E-7			
	Age	0.0000016995	6.616757E-8	4,006.14	<.0001	
	Age^2	0.0000000210	2.6339174E-9	101.84	<.0001	.365
	Age^3	0.0000000001	9.32484E-11	1.81	.1791	
	Sex	−0.0000026562	1.4313197E-6	31.32	<.0001	.003
	Age*sex	0.0000000519	6.2552832E-8	0.01	.9101	
	Age^2*sex	−0.0000000104	3.4660044E-9	9.00	.0027	.001
CRT ISD	Intercept	2.499384309	0.00144255			
	Age	0.002044210	0.00011700	1,940.54	<.0001	
	Age^2	0.000030828	0.00000419	86.38	<.0001	.218
	Age^3	−0.000000470	0.00000018	0.68	.4111	
	Sex	−0.017092024	0.00217163	159.80	<.0001	.017
	Age*sex	−0.000374711	0.00017699	0.08	.7744	
	Age^2*sex	−0.000009681	0.00000613	0.10	.7575	
	Age^3*sex	0.000000654	0.00000028	5.62	.0178	.001

Note. *F* tests are based on Type I sums of squares, so the significance of each effect is conditional on all prior effects in the model. The values of eta squared are for age, sex, and the Age × Sex interaction. Sex is coded so that women are the reference group. SRT = simple RT; ISD = intrasubject standard deviation; CRT = choice RT; age^2 = age squared; age^3 = age cubed.

Figure 3 shows the plotting of the coefficient of variation for simple RT and choice RT by age and sex. The upper panels plot the mean and standard errors in 2-year age groups. As in Figure 1, the male and female plots are slightly offset. The lower panels show the predicted values and confidence intervals for the fitted models. To aid comparison, we drew all four plots to the same scale. Parameter estimates for the models are given in Table 2. Quite different patterns are evident: Simple RT has a curvilinear relationship to age with no significant gender difference, whereas choice RT shows only a slight, and mainly linear, increase with age but a clear gender difference, reflecting the gender difference in ISD. The decrease in simple RT ISD suggested by Figure 1 and Figure 2 is more evident here.

Initial modeling of simple RT ISD controlling for the mean resulted in a model with quadratic terms in simple RT mean and age, with no significant effect of sex. To check the functional form of the relationships with RT mean and age, we compared the results with those from a semiparametric model that included spline smoothers for each and used the generalized cross-validation function to determine the degree of smoothing. This suggested that a more complex polynomial was needed and led us to fit a model with quintic terms in both age and simple RT mean. Figure 4 shows the partial regression plots from the semiparametric models with the partial fit from the parametric model overlaid as a dashed line. Because the semiparametric models are exploratory, interpretation of the results should be conservative. Nonetheless, some of the main features can be discerned. The relationship between the mean and the variability, controlling for age, is linear throughout the main range. The departure from linearity is confined to the top 1% of the distribution. The age pattern of simple RT variability, controlling for the mean, broadly agrees with results from the analysis of the coefficient of variation and from the earlier models in which the mean was not controlled. The semiparametric analysis and the higher degree polynomial suggest a flatter trajectory between the initial decline observed in people in their 20s until the beginning of the steep increase observed around 60 years of age. The estimates from the parametric model are given in Table 3.

Table 4 shows the results of modeling choice RT ISD controlling for the mean and number of errors. When cubic effects of both variables were included in the model, age was no longer significant. Introducing linear and quadratic terms attenuated the effect of age, but it remained significant until the cubic terms were entered. The gender difference was attenuated but remained significant. As with simple RT ISD, the results of this model were compared with a semiparametric model using spline smoothers for the mean and the number of errors. The right hand panels of Figure 4 show the partial regression fits from these models, again with a dashed line indicating the corresponding fit from the parametric model. There is good agreement of the main range of the data. The relationship between the mean and the variability is linear in the middle 98% of the range. As with simple RT ISD, the relationship to the mean appears different for the top 1% of the distribution. There is also some suggestion of a departure from linearity in the bottom 1%.

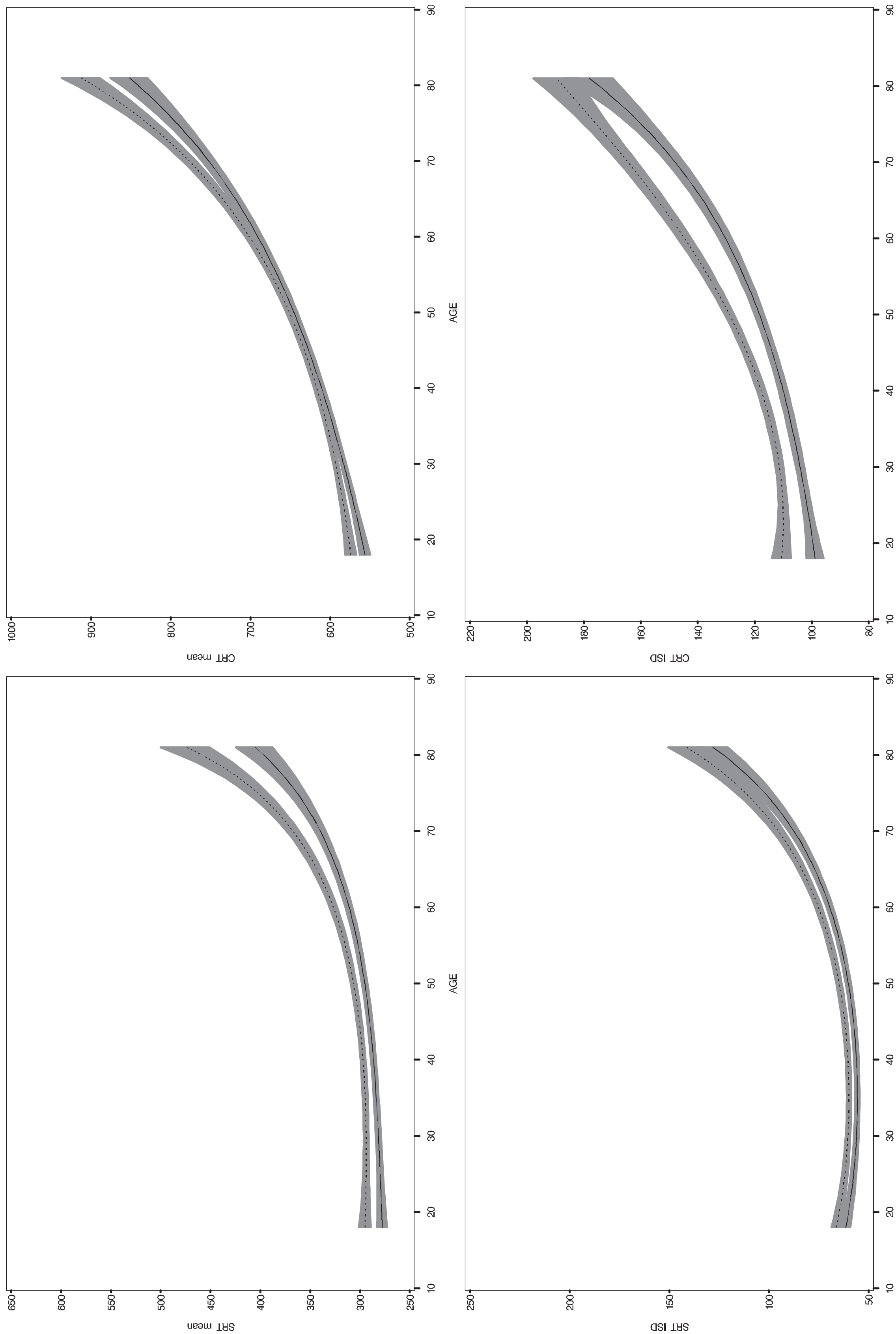

Figure 2. Predicted values and confidence intervals for reaction time measures (in milliseconds) by age and sex. Solid line = men; dashed line = women. SRT = simple reaction time; CRT = choice reaction time; ISD = intrasubject standard deviation.

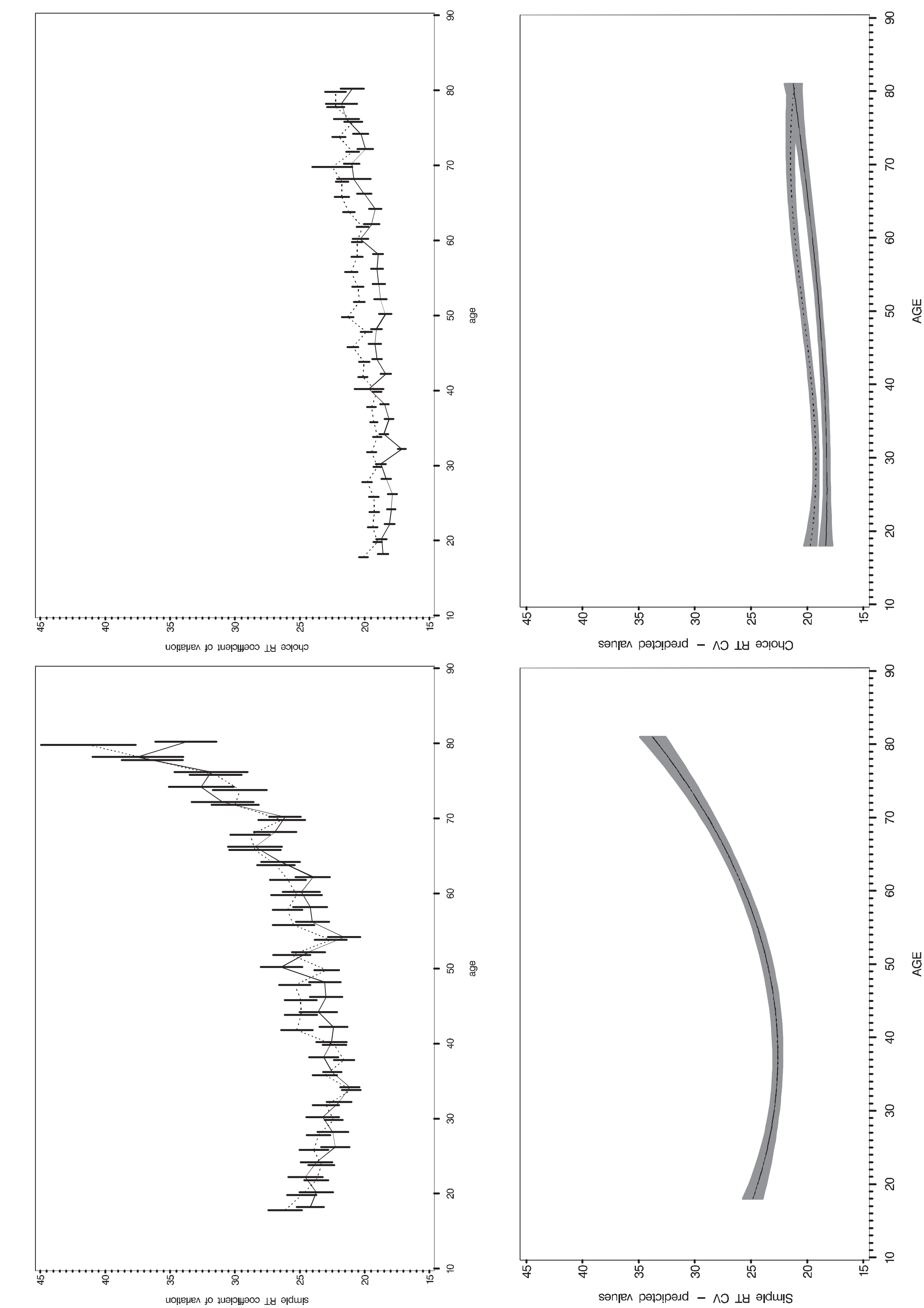

Figure 3. Simple and choice reaction time coefficient of variation: means and standard errors and predicted values and confidence intervals. Solid line = men; dashed line = women. SRT = simple reaction time; CRT = choice reaction time; CV = coefficient of variation.

Table 2
Results for Models of Simple and Choice Reaction Time Coefficient of Variation

Outcome	Parameter	Estimate	*SE*	*F*	$p > F$
SRT CV	Intercept	22.94380512	0.22589922		
	Age	0.08742994	0.00992201	165.55	<.0001
	Age^2	0.00587227	0.00054712	115.20	<.0001
CRT CV	Intercept	19.94761827	0.11426463		
	Age	0.07526313	0.00926781	165.86	<.0001
	Age^2	0.00091284	0.00033216	5.05	.0247
	Age^3	−0.00005861	0.00001443	9.03	.0027
	Male	−1.27156240	0.17201528	96.79	<.0001
	Age*male	−0.03153821	0.01401967	0.03	.8518
	Age^2*male	0.00008442	0.00048576	3.31	.0689
	Age^3*male	0.00005238	0.00002188	5.73	.0167

Note. SRT = simple reaction time; CV = coefficient of variation; CRT = choice reaction time; age^2 = age squared; age^3 = age cubed.

Variability increases with the number of errors made. The biggest difference (~11 ms.) is between those who make no errors and those who make one. Thereafter, the increase with each additional error diminishes.

Discussion

We will summarize and discuss our findings under three headings: patterns of aging for the four RT measures, sex differences, and the relationship between the mean and variability of the two RT measures.

Patterns of Aging

The most notable result is that simple and four-choice RT age differently. Both age nonlinearly, but aging for simple RT is more markedly nonlinear. The simple RT mean barely increases until people reach approximately 50 years of age, whereas the choice

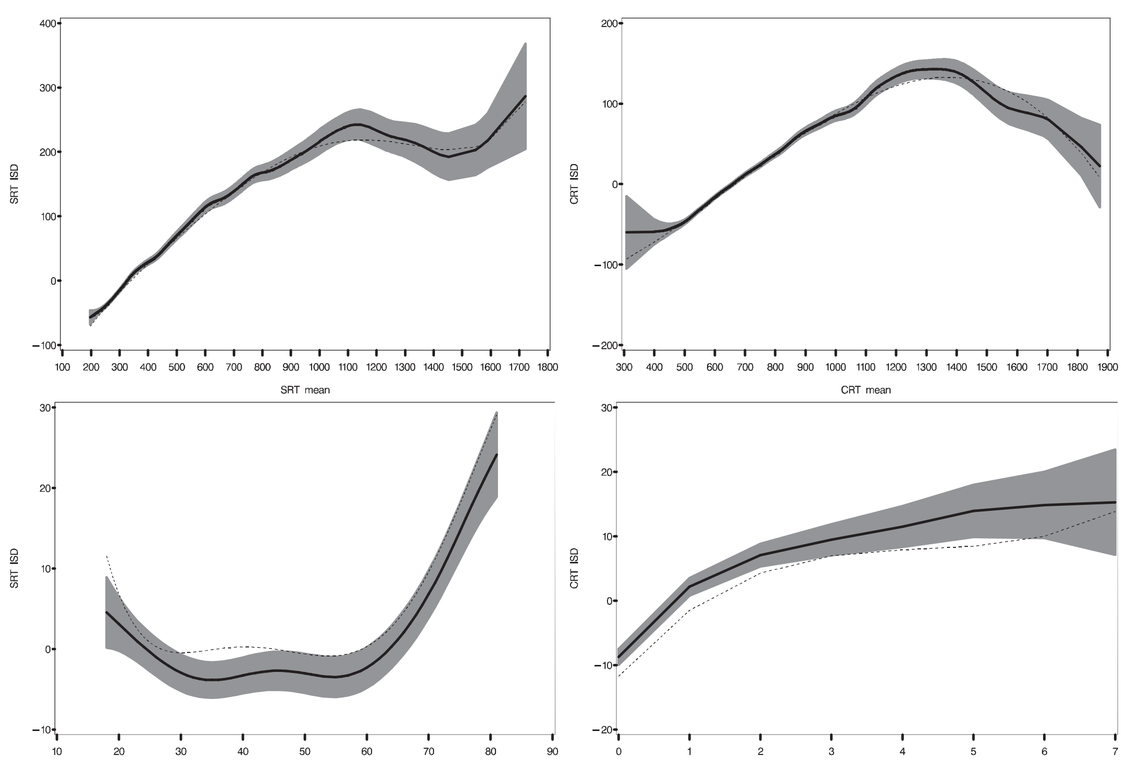

Figure 4. Partial regression plots for models of simple and choice reaction time intrasubject standard deviation (ISD) controlling for the mean. Solid line = spline smoother; shaded area = confidence band; dashed line = polynomial regression predicted values. SRT =simple reaction time; CRT = choice reaction time. Reaction time variations are given in milliseconds.

Table 3
Estimates From Model of Simple Reaction Time (SRT) ISD Controlling for the Mean

Parameter	Estimate	*SE*	*F*	*p* > *F*
Intercept	−129.3588217	30.31587857		
SRT mean	1.0019975	0.26937844	6,879.72	<.0001
SRT mean^2	−0.0016852	0.00086651	379.81	<.0001
SRT mean^3	0.0000024	0.00000125	1.44	.2305
SRT mean^4	−1.7272E-9	8.2259E-10	2.93	.0869
SRT mean^5	4.4484E-13	1.978 1E-13	5.80	.0161
Age	−0.1048165	0.10114728	16.95	<.0001
Age^2	−0.0081782	0.00840201	70.51	<.0001
Age^3	0.0007556	0.00030182	3.25	.0716
Age^4	0.0000341	0.00001158	1.13	.2885
Age^5	−0.0000008	0.00000029	7.80	.0052

Note. ISD = intrasubject standard deviation; SRT mean^2 = SRT mean squared; SRT mean^3 = SRT mean cubed; SRT mean^4 = SRT mean quadrupled; SRT mean = SRT mean quintupled; age^2 = age squared; age^3 = age cubed; age^4 = age quadrupled; age^5 = age quintupled.

RT mean increases throughout the adult age range. The variabilities show approximately the same pattern as their corresponding means, except for two differences: Simple RT variability decreases in early adulthood, whereas the mean does not; and for women, choice RT mean increases more rapidly from age 70 than the variability. These differences are mirrored in the different age patterns of the coefficient of variation.

The difference in overall pattern is intriguing. It requires explanation in the context of the processing speed hypothesis of cognitive aging and for critics of the hypothesis, who maintain instead that speed is a stable nonaging source of individual differences. Part of such an explanation might lie in the different cognitive loads that the two tasks involve. For this, there is relevant evidence from the West of Scotland Twenty-07 study conducted by the Social and Public Health Sciences Unit of the Medical Research Council of the United Kingdom. The same RT procedure was used in the study, but researchers also administered Part 1 of the Alice Heim 4 Test of General Intelligence to 900 community-based adults whose mean age was approximately 56. Choice RT was more highly correlated with the intelligence test score than with simple RT (−.49 vs. −.31; Deary et al., 2001). The relationship underlying these correlations was approximately linear for choice RT but complex and nonlinear for simple RT, suggesting that the latter has little cognitive load at above-average ability levels (Der & Deary, 2003). However, because the sample was approximately 56 years old, this population is close to the point at which age differences in simple RT mean increase. Similar data are needed from older and younger subjects.

Another, possibly complementary, explanation could be that the two RT tasks require different practice periods to reach optimal speed. Rabbitt (1993) showed that older and less able subjects need more practice to reach their peak performance. In comparison with the tasks on which Rabbitt's results are based, the four-choice RT task used in our study would be classified as very easy and, therefore, one in which peak performance could be quickly attained. Moreover, the need for more practice could equally be regarded as part of normal aging and thus appropriately reflected in the resulting RTs.

These two explanations could be regarded as components of a more general phenomenon—the age–task complexity effect. However, Salthouse (1992) concluded that this effect was primarily due to the demands on working memory, which are relatively low in both of the RT tasks used here.

One further piece of evidence relevant to the question of whether simple or choice RT better reflects cognitive aging is their relative predictive validity. The 900 subjects of the Twenty-07 study mentioned above were followed up with respect to mortality to age 70. Choice RT was found to be a better predictor of mortality than simple RT and even better than the score on the Alice Heim 4 Test of General Intelligence (Deary & Der, 2005a).

Table 4
Estimates From Model of Choice Reaction Time (CRT) ISD Controlling for the Mean

Parameter	Estimate	*SE*	*F*	*p* > *F*
Intercept	−6.57045976	16.18467401		
CRT mean	0.03733208	0.05705502	6696.96	<.0001
CRT mean^2	0.00037165	0.00006386	271.35	<.0001
CRT mean^3	−0.00000019	0.00000002	93.88	<.0001
Errors	12.92143070	1.29162500	242.61	<.0001
Errors^2	−2.90978654	0.61324601	55.02	<.0001
Errors^3	0.22658497	0.07141164	8.55	.0035
Sex	−7.94190010	0.80107272	98.29	<.0001

Note. Sex is coded so that women are the reference group. ISD = intrasubject standard deviation; CRT mean^2 = CRT mean squared; CRT mean^3 = CRT mean cubed; errors^2 = errors squared; errors^3 = errors tripled.

However, we should not forget that simple and choice RTs are highly correlated. In this study, the means are correlated at $\rho = .67$, implying that 45% of the variance in choice RT can be explained by the simpler measure. Simple RT and the reasons behind its increase with age remain questions worth studying.

The decrease in simple RT variability suggested in Figure 2 (and more evident after controlling for the mean) has been noted previously. Pierson and Montoye (1958) studied simple RT in 400 males between the ages of 8 and 83. Using frequency of modal response as their measure of consistency, Pierson and Montoye concluded that "there is an increase in consistency of response with age from eight years until about 30, after which a decline is evident. Although an individual is capable of the fastest response at about age 20, he is most consistent about 10 years later" (p. 419). In a more recent study of 291 individuals ranging in age from 6 to 89, Li et al. (2004) specifically focused on variability, which they referred to as robustness. They concluded that "maximum processing speed, processing robustness, and fluid intelligence were achieved by individuals in their mid 20s" (p. 159). However, their measure of robustness was a composite, and their numbers were rather small for such a large age range. If, as Rabbitt and others have suggested, RT variability plays an important role in cognitive aging, it may also do so during the developmental phase and thus warrant further study in that context.

Sex Differences

Sex differences were found for each of the four measures examined, with the strongest and most robust being that for choice RT variability. There was a suggestion of this effect in our analysis of data from the Twenty-07 study (Deary & Der, 2005b), but otherwise we are not aware of its having been previously reported. It was still significant, although attenuated somewhat, in the analysis of variability controlling for the mean even when the effect of age was no longer significant.

One possible explanation is that it is due to some sex difference in the speed–accuracy tradeoff. We have examined this in two ways. In the analysis of variability controlling for the mean and number of errors, we tested for interactions between sex and error rate, but none were significant. We also analyzed only those who made no errors (results not shown), and the sex difference remained.

A further possibility, suggested by Reimers and Maylor (in press), is that the difference may reflect systematic sex difference in performance across a block of trials. Their study included a 12-trial two-choice RT task in which participants pressed a key to indicate whether a face presented briefly on screen was male or female. They analyzed data from 5,137 individuals. They also found RT in women to be more variable than that in men and the effect remained in an analysis of the coefficient of variation. In addition, they found that women were slower, but more accurate, than men in the first two trials and slightly faster with the same error rate in Trials 3 to 12. Excluding the first two trials eliminated the sex difference in variability. They concluded that trial-to-trial shifts in the speed–accuracy tradeoff could explain the sex difference in variability. Although their design may have exaggerated trial-to-trial effects by not having a practice phase, their hypothesis merits further research. Not having the data for individual trials, we were not able to test it with the HALS data.

The sex differences for choice RT mean are the weakest and most variable across the age range. Given that choice RT is more widely studied than simple RT and RT mean more so than variability, it is easy to see how other studies with smaller samples and incomplete age coverage could yield equivocal results.

Both simple and choice RT mean increases more rapidly for women at older ages. This could partly reflect sex differences in survival to these ages, whereby the men still alive and eligible for sampling would represent a healthier subset. However, for choice RT, the simultaneous narrowing of variability would not fit with this explanation.

Reaction Time Variability

When the RT mean and variability are combined in the coefficient of variation, the age patterns are again different for simple and choice RT. Choice RT is comparatively flat, with only a slight, mainly linear, increase with age and the sex differences in choice RT variability is still evident in the coefficient of variation. In contrast, the simple RT coefficient of variation is markedly nonlinear and without significant sex differences.

The results are extended in the analyses of RT variability controlling for RT mean. When choice RT variability is adjusted for its mean and the number of errors, there is no longer a significant age effect. Cubic terms in choice RT mean and number of errors are needed to reduce the age effect to nonsignificance. A positive relationship between the number of errors and choice RT variability is expected even though the results are based only on correct responses, because the responses immediately following an error are known to be slower (Rabbitt, 1969).

Simple RT variability retains a significant relationship to age after one controls for its mean. The decrease in simple RT variability in early adulthood, discussed above, was more evident in the analyses in which the mean was controlled for, suggesting that studies with smaller sample sizes, such as the sample size in the study by Li et al. (2004), may be able to increase their power by controlling for mean RT.

Strengths of the Study

This study has a number of strengths. Foremost among these is the quality of the sample, which is very large, is representative of the population, and covers the whole adult age range. The RT task is of a standard format and uses a device designed and built especially for the study. Both means and variabilities were collected for simple and choice RTs. The Box-Cox procedure was used to find a data transformation that was optimal for the analysis.

Limitations of the Study

The main limitation of the study is that the device used to record RTs did not store the results of the individual trials. Their absence precludes direct testing of some principal hypotheses. One is the suggestion made by Rabbitt et al. (2001) that the slowing of RT with age is mainly due to an increase in the number of atypically slow responses. Another is the hypothesis of Reimers and Maylor (in press) that was mentioned earlier.

Because the interviews were conducted in the respondents' homes, the test conditions were not as rigorously controlled as test conditions in a laboratory setting. The data are cross-sectional and, therefore, inferences from the age differences that they exhibit to the aging process are subject to potential confounding, notably by period and cohort effects as well as by survivorship bias at the older ages.

Conclusions

This study provides a detailed description of the aging patterns of simple and four-choice RT means and variabilities. The patterns differ markedly between simple and choice RTs even though the two are highly correlated. Choice RT slows and becomes more variable throughout adulthood, whereas simple RT barely changes until people are approximately 50 years old. If speed, as measured by RTs, deserves "special status in the context of cognitive aging" (Verhaeghen & Salthouse, 1997, p. 246), then these results need to be understood.

The relationship between RT mean and variability, whether represented in the coefficient of variation or in analyses of variability controlling for the mean, also has different age patterning for simple and choice RTs.

Sex differences are demonstrated for each of the measures, with the most consistent and robust being that for choice RT variability. This effect remained significant, even in analyses that accounted entirely for age differences. This is a novel finding that awaits a definitive explanation.

In 1987, Nettelbeck called for studies of large representative samples in the field of processing speed measures. The original report of the HALS study (Cox et al., 1987) was published in the same year. By revisiting and reanalyzing the data, we have both confirmed and elaborated on the original descriptive results in exactly the way Nettelbeck envisaged.

References

Anstey, K. J. (1999). Sensorimotor variables and forced expiratory volume as correlates of speed, accuracy, and variability in reaction time performance in late adulthood. *Aging, Neuropsychology, and Cognition, 6,* 84–89.

Blaxter, M. (1987). Sample and data collection. In B. D. Cox (Ed.), *The Health and Lifestyle Survey* (pp. 1–3). London: Health Promotion Research Trust.

Botwinick, J., & Storandt, M. (1974). Cardiovascular status, depressive affect, and other factors in reaction time. *Journal of Gerontology, 29,* 543–548.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B, 26,* 211–252.

Cattell, R. B. (1950) Culture Fair Intelligence Test: A measure of "g." Savoy, IL: Institute for Personality and Ability Testing.

Cox, B. D., Blaxter, M., Buckle, A. L. J., Fenner, N. P., Golding, J. F., Gore, M., et al. (1987). *The Health and Lifestyle Survey.* London: Health Promotion Research Trust.

Crook, T. H., & West, R. L. (1990). Name recall performance across the adult life-span. *British Journal of Psychology, 81,* 335–349.

Deary, I. J., & Der, G. (2005a). Reaction time explains IQ's association with death. *Psychological Science, 16,* 64–69.

Deary, I. J., & Der, G. (2005b). Reaction time, age and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging Neuroscience and Cognition, 12,* 187–215.

Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence, 29,* 389–399.

Der, G., & Deary, I. J. (2003). IQ, reaction time and the differentiation hypothesis. *Intelligence, 31,* 491–503.

Fozard, J. L., Vercruyssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore Longitudinal Study of Aging. *Journal of Gerontology: Psychological Sciences, 49,* 179–189.

Fuentes, K., Hunter, M. A., Strauss, E., & Hultsch, D. F. (2001). Intraindividual variability in cognitive performance in persons with chronic fatigue syndrome. *Clinical Neuropsychologist, 15,* 210–227.

Harkins, S. W., Nowlin, J. B., Ramm, D., & Schroeder, S. (1974). Effects of age, sex, and time-on-watch on a brief continuous performance task. In E. Palmore (Ed.), *Normal aging II: Reports from the Duke longitudinal studies, 1970–1973*. Durham, NC: Duke University Press.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models.* London: Chapman and Hall.

Heim, A. W. (1970). *Manual for the AH4 group test of general intelligence.* Windsor, England: National Foundation for Educational Research.

Hofer, S. M., & Sliwinski, M. J. (2001). Understanding ageing: An evaluation of research designs for assessing the interdependence of ageing-related changes. *Gerontology, 47,* 341–352.

Hultsch, D. F., & MacDonald, S. W. S. (2004). Intraindividual variability in performance as a theoretical window onto cognitive aging. In R. A. Dixon, L. Backman, & L.-G. Nilsson (Eds.), *New frontiers in cognitive aging.* Oxford, England: Oxford University Press.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology: Psychological Sciences, 57B,* 101–115.

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology, 14,* 588–598.

Huppert, F. A. (1987). Cognitive Function. In *Health and Lifestyle Survey.* London: Health Promotion Research Trust.

Johnson, R. C., McClearn, G. E., Yuen, S., Nagoshi, C. T., Ahern, F. M., & Cole, R. E. (1985). Gallon's data a century later. *American Psychologist, 40,* 875–892.

Koga, Y., & Morant, G. M. (1923). On the degree of association between reaction times in the case of different senses. *Biometrika, 15*(3/4), 346–372.

Li, S-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science, 15,* 155–163.

Madden, D. J. (2001). Speed and timing of behavioural processes. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed.). San Diego, CA: Academic Press.

Mathey, F. J. (1976). Psychomotor performance and reaction speed in old age. In D. Kuhn (Series Ed.) & H. Thomae (Vol. Ed.), *Contributions to human development: Vol. 3. Patterns of aging: Findings from the Bonn Longitudinal Study of Aging* (pp. 36–50). Basel, Switzerland: Karger.

Medical Research Council, Social and Public Health Sciences Unit. (2005). *West of Scotland Twenty-07 Study: Health in the community*. Retrieved February 21, 2006, from http://www.msoc-mrc.gla.ac.uk/Twenty-07/Twenty-07_MAIN.html

Nettelbeck, T. (1987). Inspection time and intelligence. In P. A. Vernon (Ed.), *Speed of information processing and intelligence.* Norwood, NJ: Ablex.

Pearson, K. (1924). *The life letters and labours of Francis Galton: Volume II. Researches of middle life*. Cambridge, England: Cambridge University Press.

Pierson, W. R., & Montoye, H. J. (1958). Movement time, reaction time, and age. *Journal of Gerontology, 13,* 418–421.

Rabbitt, P. (1969). Psychology refractory delay and response-stimulus interval duration in serial, choice-response tasks. In W. Koster (Ed.), *Attention and performance II* (pp. 195–219). Amsterdam: North-Holland.

Rabbitt, P. (1993). Crystal quest: A search for the basis of maintenance of practised skills into old age. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: selection, awareness, and control. A tribute to Donald Broadbent.* Oxford, England: Clarendon Press.

Rabbitt, P., Osman, P., Moore, B., & Stollery, B. (2001). There are stable individual differences in performance variability, both from moment to moment and from day to day. *Quarterly Journal of Experimental Psychology, 544,* 981–1003.

Reimers, S., & Maylor, E. A. (in press). Gender effects on response time variability and trial-to-trial performance: Reply to Deary and Der (in press). *Aging, Neuropsychology, and Cognition*.

Salthouse, T. A. (1992). Why do adult age-differences increase with task complexity? *Developmental Psychology, 28,* 905–918.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403–428.

Sliwinski, M., & Buschke, H. (1999). Cross-sectional and longitudinal relationships among age, cognition, and processing speed. *Psychology and Aging, 14,* 18–33.

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122,* 231–249.

Received September 16, 2004
Revision received August 8, 2005
Accepted September 9, 2005 ■