J. Sleep Res. (2006) **15,** 376–385

Age and individual variability in performance during sleep restriction

PAUL D. BLIESE, NANCY J. WESENSTEN and THOMAS J. BALKIN
Department of Behavioral Biology, Division of Psychiatry and Neuroscience, Walter Reed Army Institute of Research, Silver Spring, MD, USA

Accepted in revised form 10 September 2006; received 30 November 2005

SUMMARY The effect of sleep loss on reaction time (RT) performance varies as a function of age, with RTs of older subjects typically showing less decrement (relative to rested baseline) than those of younger subjects. In the current paper, we examined the nature of this relationship in a 7-day sleep restriction study. The number of repeated measures made it possible to model both intra-individual trajectories over days and individual differences in these trajectories. Results revealed (a) consistent individual differences in RT patterns over time after controlling for experimental design effects; (b) less cumulative RT decline among older individuals regardless of the degree of sleep restriction; and (c) consistent individual variability in performance patterns even after accounting for the effects of age.

KEYWORDS age, growth modeling, performance, sleep restriction

The effect of sleep loss on reaction time (RT) performance varies as a function of age, with RTs of older subjects typically showing less decrement (relative to baseline) than those of younger subjects (e.g. Bonnet, 1989; Bonnet and Arand, 1989; Bonnet and Rosa, 1987; Philip *et al*., 1999, 2004; Stenuit and Kerkhofs, 2005). In most relevant studies, the effects of age have been investigated in short-duration experiments generally involving 1 or 2 days of sleep deprivation or 3 days of sleep restriction. These studies produce relatively few repeated measures upon which to examine individual differences related to age. In addition, prior studies usually employ analytic techniques such as repeated measures ANOVA. In repeated measures ANOVA, age effects over time may not be readily discernable for two reasons. First, the technique assumes that the groups of subjects will respond similarly to experimental conditions and is thereby not designed to model individual differences among subjects (Van Dongen *et al*., 2003). Second, ANOVA-based analytic techniques often categorize subjects into 'older' and 'younger' groups potentially reducing individual difference effects associated with age.

Correspondence: Paul D. Bliese, PhD, Division of Psychiatry and Neuroscience, Walter Reed Army Institute of Research, 503 Robert Grant Avenue, Room # 2A26, Silver Spring, MD 20910, USA. Tel.: +1-301-319-9350; fax: +1-301-319-9979; e-mail: paul.bliese@us.army.mil

In sleep research, analytic approaches that consider individual differences are becoming increasingly important because studies have identified consistent and reliable individual differences in performance during sleep loss conditions (Van Dongen *et al*., 2003, 2004a,b). For instance, after controlling for experimental design effects, Van Dongen *et al*. (2003) estimated subject-specific rates of change in psychomotor vigilance test (PVT) lapses across 14 days, and identified consistent individual differences in rates of change over time. In another study, Van Dongen *et al*. (2004a) modeled individual differences in participant performance across three separate 36-h total sleep deprivation sessions and found trait-like differences in performance across the three sessions as evidenced by intra-class correlation (ICC) values ranging from 0.675 to 0.922. These findings suggest that: (a) some individuals are more vulnerable to the effects of sleep loss than others, and (b) there is consistency in the extent to which individual performance is impacted by sleep loss across time.

Two approaches have been used to examine how individual traits or characteristics such as age relate to performance. One approach is to examine the degree to which individual traits predict performance across situations; that is, the degree to which individuals replicate performance across sessions. For instance, Van Dongen *et al*. (2004a) used standard regression

techniques to identify predictors of individuals' average performance across three sessions each involving 36 h of total sleep deprivation. These predictors included demographic variables such as age, sex, and body mass index as well as psychosocial traits such as extraversion, neuroticism, and trait anxiety. Analyses revealed that only 'overall sleepiness,' 'impatience,' and perceptions of stressful conditions were predictive of individual differences in performance – age was not related to average levels of individual vulnerability across the multiple sessions.

The second approach is to examine individual difference profiles of performance within (rather than across) long-duration sessions. Studies involving more than three periods within a single session (in this case, days) are particularly valuable because they provide flexibility in estimating performance trajectories over time. Performance trajectories can be examined both in terms of average trajectories for the entire sample and in terms of individual-specific trajectories (Olofsen *et al*., 2004; Singer and Willett, 2003; Van Dongen *et al*., 2003). For example, the 14-day Van Dongen *et al*. (2003) study provided the opportunity to describe and contrast the average pattern of repeated measures in the 8-, 6- and 4-h sleep restriction conditions. The study also estimated subject-specific rates of change for performance on the PVT, and identified that subjects' sleep history during the 5 days prior to the experiment was related to their unique individual trajectories.

The examination of intra-individual performance trajectories in long-duration sessions allows questions of individual differences to be framed not only in terms of *extent* (i.e. amount of resulting degradation) but also in terms of *pattern* (e.g. linear rate of decline and quadratic rate of decline). Thus, analyzing data collected over multiple days of sleep deprivation or sleep restriction can facilitate analyses of individual traits and characteristics such as age and allow greater elucidation of their relationship to performance (see Olofsen *et al*., 2004; Van Dongen *et al*., 2003).

In the present paper, we extended the work of Van Dongen and others by using discontinuous growth modeling (DGM) (Singer and Willett, 2003) to examine both intra-individual patterns of change and individual differences in these patterns across multiple days of sleep restriction. The DGM provides the opportunity to describe the intra-individual patterns of change in terms of three distinct parameters – an experimental slope, a recovery transition parameter, and a recovery slope. It was hypothesized that reliable differences in responses to sleep restriction (both in terms of overall performance and trend patterns) would be evident after controlling for differences engendered by experimental design factors. Analyses examine the extent to which participant's age explains differences in intra-individual trend patterns, and thereby helps explain individual differences in vulnerability to sleep restriction.

The analyses reported here were performed on a subset of data collected in a larger study analyzed and published as a US Department of Transportation Report (Balkin *et al*., 2000; see also Balkin *et al*., 2004; Belenky *et al*., 2003).

METHODS

Subjects

Subjects were 65 healthy, normal non-smoking adults (16 women, age 24–55, mean = 43 years; and 50 men, age 24–62, mean = 37 years) holding valid Commercial Motor Vehicle (CMV) drivers' licenses. Reported daily caffeine consumption was no more than 300–400 mg per day. Volunteers were medication-free (including over-the-counter medications) beginning 48 h prior to the study (exception: women continued birth control medication).[1]

Design

Volunteers spent 14 days in-residence. The first 3 days were adaptation/training (T1, T2) and then baseline (B); for T1, T2, and B, volunteers were in bed from 23:00–07:00 hours (8 h required time in bed, TIB). Days 4–10 constituted the experimental phase (E1–E7) in which volunteers were assigned to one of four night-time sleep conditions [9 h required TIB (22:00–07:00 hours), 7 h required TIB (24:00–07:00 hours), 5 h required TIB (02:00–07:00 hours), or 3 h required TIB (04:00–07:00 hours)]. Days 11–13 constituted the recovery phase (R1–R3) during which all volunteers were given 8 h TIB (23:00–07:00 hours) per night. On day 14, volunteers were released from the study (no data were collected). Throughout all phases of the study, lights were switched on at 07:00 hours. Volunteers were not permitted any other time in bed except as required by periodic sleep latency tests.

Psychomotor vigilance test

Volunteers attended to an LED timer display and pressed the response button as quickly as possible when the stimulus appeared. The stimulus consisted of the LED timer turning on and incrementing from zero at 1-ms intervals. A button press stopped the timer from incrementing and displayed response latency for 0.5 s. The display then turned off for the remainder of foreperiod preceding the next stimulus. Foreperiods varied randomly from 2 to 10 s. The dependent measure used in analyses reported in the present paper was mean RT per day collapsed across time of day. RTs above 500 ms were excluded from the calculation of mean RT.

Polysomnography

Polysomnographic (PSG) measures [EEG (C3 and C4); EOG (outer canthi of each eye), EMG (mental/submental)], and EKG (from just below left and right clavicle) were recorded continuously throughout the study using Medilog 9000-II magnetic cassette recorders (Oxford Instruments, Largo, FL, USA). PSG data from night-time sleep periods were scored

[1]The Belenky *et al*. (2003) sample contained 66 subjects, but one individual in the 7-h TIB condition lacked reliable objective sleep data and was therefore excluded.

offline in accordance with Rechtschaffen and Kales (1968) criteria by six technicians whose inter-rater reliabilities were at least 85% compared with the scoring of a diplomate of the American Board of Sleep Medicine (TJB). The independent measure used in analyses reported here was average total sleep time per night (TST) during the experimental phase (average of nightly minutes spent in all sleep stages).

Procedure

Volunteers reported to the laboratory at 10:00 hours on the day prior to T1. Verbal and written descriptions of study procedures were given, and electrodes for ambulatory polysomnography were applied. Volunteers then underwent training on the various performance tasks. Throughout the study, meals were served at 08:30, 12:30, and 17:30 hours, with snacks and beverages available *ad libitum* between performance tests. Volunteers did not use/consume nicotine or caffeine-containing products during the study; random urine drug screens verified compliance. Use of medications during the study (e.g. acetaminophen for headache) was allowed at the discretion of the attending physician. For all women enrolled in the study, serum pregnancy tests performed at the beginning of the study were negative.

Days T1 and T2 consisted of training on performance tests and familiarization with study procedures. Baseline testing occurred on B1, and testing continued for the duration of the study (E1–E7, R1–R3). Following a final 8-h sleep period in the laboratory, electrodes were removed shortly after awakening, and volunteers were debriefed and released from the study.

A detailed description of study methods and procedures can be found in Balkin *et al*. (2000).

Analytic strategy

Sleep and performance data used in analyses

Sleep was objectively recorded during the experiment. Consequently, each volunteer's actual TST was used for analyses rather than TIB experimental condition. TST was calculated using each volunteer's average amount of sleep across the seven experimental nights. TST was related to the TIB manipulation. The 3-h TIB group slept an average of 171.89 min (SD = 5.79) each night of the experiment. The 5-h TIB group slept 279.87 min (SD = 7.81); the 7-h group slept 376.90 min (SD = 19.71), and the 9-h group slept 477.39 min (SD = 20.36).

The PVT was administered four times per day. However, because time-of-day effects were not the focus of this analysis, these four PVTs were collapsed into a single daily measure (one each for B, E1–E7, R1–R3; resulting in 11 data points per volunteer).

Random coefficient modeling

Analyses were based on random coefficient modeling (RCM) (Hox, 2002; Pinheiro and Bates, 2000; Raudenbush and Bryk, 2002), also referred to as mixed effects modeling (Littell *et al*., 1996). The transition between the experimental phase and the recovery phase resulted in abrupt performance changes (as seen in PVT data presented in Belenky *et al*., 2003) which are not appropriately captured in smooth, continuous parameter models (Singer and Willett, 2003). Thus, an RCM growth model for discontinuous or nonlinear change (Singer and Willett, 2003) was selected. All analyses were conducted using the open-source platform R (R Development Core Team, 2005) and the nonlinear and linear mixed effect model (NLME) package for R (Pinheiro and Bates, 2000).

A growth modeling strategy similar to that described by Bliese and Ployhart (2002) was used in the analyses. Briefly, growth modeling is divided into two major phases with several steps within each phase. Phase 1 consists of identifying and modeling the nature of the intra-individual growth trajectories over time and determining the degree to which these growth trajectories contain reliable individual differences. This phase can be conducted using three steps detailed in Table 1. A major difference between the strategy presented by Bliese and Ployhart (2002) and the strategy presented in Table 1 is that

Table 1 Discontinuous growth modeling analytic strategy

Phase	*Step*	*Description/purpose*
1	1	Model basic design elements of experiment in a random intercept model; ensure within-individual error structure (autocorrelation) is adequately modeled to avoid bias. Partition residual variance into between-individual and within-individual components, and calculate conditional intra-class correlation coefficient (ICC). ICC provides an estimate of how much total variance in RT can be explained by individual differences.
	2	Determine whether the sample-based growth patterns identified in step 1 vary significantly across individuals by testing for inter-individual differences in: (a) experimental phase slope; (b) recovery transition point; (c) recovery phase slope.
	3	Re-examine within-individual error structure (autocorrelation, heteroscedasticity) to ensure final design model is not biased.
2	1	Examine the degree to which level 2 predictors such as participant age help explain sources of variance identified in Phase 1, Step 2.

applying RCM to experimental designs requires one to build a phase 1 model reflecting the experimental design elements. Sources of variability (the focus of phase 1) are estimated after controlling for design elements.

Phase 2 consists of identifying individual factors that explain individual differences in the intra-individual growth trajectories. In this phase, the effect of age is investigated.

Potential age confounds

To help interpret subsequent results, an analysis of age differences was conducted. A regression model predicting daily RT performance at baseline revealed no significant age effect ($t = 0.35$, $P = 0.72$). This suggests older and younger participants were performing similarly at baseline. A second analysis indicated no significant differences in age in the four experiment TIB conditions ($F[3,62] = 1.56$, $P = 0.21$) suggesting age differences were equally distributed across TIB groups. These findings were mirrored in a third analysis examining the relationship between age and TST. This analysis also revealed no differences ($t = 0.29$, $P = 0.78$) suggesting that TST was unrelated to age. An analysis of gender and age revealed that older individuals tended to be male ($t = -2.05$, $P < 0.05$). Thus, in the final analyses, the effects of gender, age, and the interaction between age and gender were examined to ensure the observed findings related to age were not spurious gender effects.

RESULTS

Results are presented for each phase of the growth curve analysis and for the steps within each of these phases (see the Analytic strategy section).

Phase 1

Step 1

A null RCM model was estimated as the basis for calculating the ICC. The ICC represents the degree to which reliable individual differences in mean RT exist among participants. In experimental settings, it is common to use a null model reflecting the experimental design; thus, the ICC is conditional on the design elements.

In the null model, the design elements were included using the discontinuous growth model specification as detailed in Singer and Willett (2003). In this model, the effects of time were captured using three level 1 predictors. The first predictor (EXPER) was a vector of sequential whole numbers ranging from zero (baseline) to 10 (final recovery day). The second predictor (TRANS) was a dummy coded vector containing a value of zero for RT data collected during the experimental phase, and a value of one for RT measures collected during the recovery phase. The third predictor (RECOV) was a vector containing zeros for RTs collected during the experimental phase and sequential numbers from zero to two for RTs collected across the three recovery days. Table 2 shows the data structure for one participant in the 3-h TIB group. Notice that each RT is predicted by a unique combination of EXPER, TRANS and RECOV values.

Table 2 Data structure for one participant

SUBJECT	*RT*	*EXPER*	*TRANS*	*RECOV*
308	255.02	0	0	0
308	313.00	1	0	0
308	359.36	2	0	0
308	392.94	3	0	0
308	398.54	4	0	0
308	384.79	5	0	0
308	394.41	6	0	0
308	403.94	7	0	0
308	329.97	8	1	0
308	349.03	9	1	1
308	340.04	10	1	2

Equation 1 formally presents the level 1 model representing time in the experimental design. The notation is based upon Singer and Willett (2003) where participant is indexed by i and time by j.

$$\mathrm{RT}_{ij} = \pi_{0i} + \pi_{1i}(\mathrm{EXPER}_{ij}) + \pi_{2i}(\mathrm{TRANS}_{ij}) + \pi_{3i}(\mathrm{RECOV}_{ij}) + \varepsilon_{ij} \quad (1)$$

$$\pi_{0i} = \gamma_{00} + \zeta_{0i} \quad (2)$$

Equation 1 captures the discontinuous, intra-individual component of the experimental design related to time. The three predictors (EXPER, TRANS and RECOV) provide parameter estimates that, respectively, correspond to: (a) the experimental phase linear slope; (b) the change in RT associated with the transition from experimental to recovery phase; and (c) the recovery phase linear slope (see Singer and Willett, 2003, p. 198–199). Equation 2 specifies that each individual's intercept is a function of an overall population intercept value (γ_{00}) plus unique error for the individual (ζ_{0i}). Equation 2 serves the purpose of allowing for individual differences in mean RT performance across trials.

The experimental design, however, also contains the manipulation of TIB. Individuals were randomly assigned to a TIB condition (9, 7, 5 or 3 h) to vary nightly sleep throughout the experimental phase. The level 2 predictor of average nightly total sleep during the experimental phase (TST) presumably impacts all aspects of the level 1 model identified in Eqns 1 and 2. Formally, the expected impact of TST is specified in Eqns 3–7:

$$\mathrm{RT}_{ij} = \pi_{0i} + \pi_{1i}(\mathrm{EXPER}_{ij}) + \pi_{2i}(\mathrm{TRANS}_{ij}) + \pi_{3i}(\mathrm{RECOV}_{ij}) + \varepsilon_{ij} \quad (3)$$

$$\pi_{0i} = \gamma_{00} + \gamma_{01}(\mathrm{TST}_i) + \zeta_{0i} \quad (4)$$

Table 3 Phase 1 design-effect model

	Parameter	*SE*	*d.f.*	t-*value*	P-*value*
Fixed effects					
Intercept (ms)	284.94	15.12	644	18.85	0.00
Experimental slope (EXPER)	17.63	1.19	644	14.86	0.00
Transition to recovery (TRANS)	−76.90	6.88	644	−11.18	0.00
Recovery slope (RECOV)	−13.09	4.68	644	−2.79	0.01
Average total nightly sleep time (TST)	−4.26	2.66	63	−1.60	0.11
EXPER × TST	−2.44	0.21	644	−11.69	0.00
TRANS × TST	11.18	1.21	644	9.25	0.00
RECOV × TST	1.75	0.82	644	2.12	0.03
Variance components					
Intercept (ζ)	1460.97				
Residual (ε)	349.82				
Fit indices					
Deviance (−2 Log-likelihood)	6193.85				
AIC	6215.85				

$$\pi_{1i} = \gamma_{10} + \gamma_{11}(\mathrm{TST}_i) \quad (5)$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}(\mathrm{TST}_i) \quad (6)$$

$$\pi_{3i} = \gamma_{30} + \gamma_{31}(\mathrm{TST}_i) \quad (7)$$

Equation 4 specifies that TST is an expected predictor of the intercept (i.e. a predictor of mean differences among individuals in RT values across time). Equation 5 specifies that TST is an expected predictor of individual differences in the RT slope during the experimental phase. Equation 6 specifies that TST is an expected predictor of individual differences in RT in the transition between the experimental and recovery phase, and Eqn 7 specifies that TST is an expected predictor of individual differences in the recovery phase slope.

The terms in Eqns 4–7 can be substituted into Eqn 3 to produce a combined a single equation:

$$\begin{aligned}\mathrm{RT}_{ij} = {} & \gamma_{00} + \gamma_{10}(\mathrm{EXPER}_{ij}) + \gamma_{20}(\mathrm{TRANS}_{ij}) + \gamma_{30}(\mathrm{RECOV}_{ij}) \\ & + \gamma_{01}(\mathrm{TST}_i) + \gamma_{11}(\mathrm{EXPER}_{ij} \times \mathrm{TST}_i) + \gamma_{21}(\mathrm{TRANS}_{ij} \\ & \times \mathrm{TST}_i) + \gamma_{31}(\mathrm{RECOV}_{ij} \times \mathrm{TST}_i) + \zeta_{0i} + \varepsilon_{ij}\end{aligned} \quad (8)$$

Equation 8 represents the final form of the discontinuous linear-mixed effects model reflecting the experimental design. Estimates from the model can be used to calculate the conditional ICC. Prior to estimating the final model for the ICC, however, an examination of the within-individual error structure was conducted (Bliese and Ployhart, 2002). These analyses suggested significant lag 1 serial autocorrelation in the repeated RT measures ($\phi = 0.51$); consequently a lag 1 within-individual error structure term was included. The model estimates are provided in Table 3.

The variance components in Table 3 provide an estimate of the conditional ICC. The value is estimated as 1460.97/(1460.97 + 349.82) or 0.81 indicating that 81% of the post-design-effect variation in RT represents consistent individual differences. This value indicates a high degree of individual consistency, and is similar to the ICC values reported by Van Dongen *et al*., 2004a).[2]

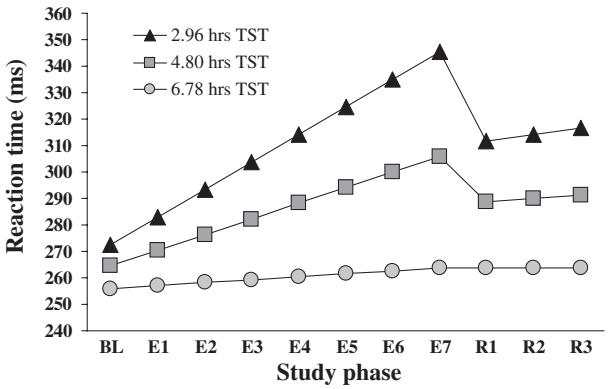

Figure 1. Predicted psychomotor vigilance test reaction time for three different individuals as a function of study phase.

Fig. 1 uses the parameter estimates from Table 3 to illustrate the experimental design effects. Estimated RT values are provided for: (a) an individual at the first quartile of the TST distribution (2.96 h average TST per night during experimental phase); (b) an individual with a median TST value (4.80 h), and an individual at the third quartile (6.78 h).

[2]An alternative model specification using TIB condition as a categorical variable (modeled as three dummy codes with the 3-h TIB condition as the referent) provides a virtually identical conditional ICC value of 0.80. This latter model, while similar to the TST model, is considerably more complex because the use of three dummy codes for the main effect and each interaction adds eight additional predictors to the model. The AIC suggests that the added complexity does not improve fit (AIC of the TIB model is 6257.573; AIC of the TST model is 6215.85). In the alternative model, TIB was coded as a categorical variable rather than modeled as an integer (3, 5, 7, or 9) because treating the variable as an integer would have assumed that the integer reflected actual sleep and that there was a constant 2-h difference across groups. These assumptions cannot be supported by the actual recorded sleep times, so the more conservative approach is to treat the variable as categorical.

Table 4 tests for slope variability in design-effect model

Model	*Random parameter*	*d.f.*	*AIC*	*−21og Lik*	*Test*	*L.ratio*	P-*value*
1	Intercept	11	6215.85	6193.85			
2	EXPER	13	6125..07	6099.07	1 versus 2	94.78	0.00
3	EXPER and TRANS	16	6098.54	6066.54	2 versus 3	32.53	0.00
4	EXPER, TRANS and RECOV	20	6102.08	6062.08	3 versus 4	4.45	0.35

The significant interactions between TST and each time parameter (EXPER, TRANS, and RECOV) are illustrated in the figure. During the experimental phase, low TST was associated with an increase in RT across days (EXPER × TST). At the transition from experimental to recovery phase, low TST was associated with a larger relative improvement in RT (TRANS × TST). During the recovery phase, low TST was associated with a slight increase in RT across days (RECOV × TST).

Step 2

Step 2 of Phase 1 involves determining the degree to which reliable inter-individual differences exist after design effects have been modeled. This step sets the stage for examining how individual difference variables such as age are related to RT over time. The conditional ICC of 0.81 calculated in step 1 establishes that there are significant individual differences in overall RT performance. Thus, the focus of step 2 is to determine whether there are reliable differences remaining in the slope and transition parameters related to time.

To examine parameter variability, −2 log-likelihood values were contrasted in models where parameters were fixed versus models that allowed parameters to vary for each participant (Bliese and Ployhart, 2002; Pinheiro and Bates, 2000; Singer and Willett, 2003). To test for variability in the three slope parameters, four models were contrasted. The first model is represented by Eqn 8 and restricts all three slope parameters to be equal across respondents. The second model allowed individual slopes to vary for the experimental phase slope (EXPER). The third model allowed for variability in both the EXPER and the transition (TRANS). The fourth model allowed for individual variability in all three parameters. More formally, the contrasts can be illustrated using Eqns 9–11. The first contrast involves using a model with Eqn 9 rather than Eqn 5 in the combined model (Eqn 8). The −2 log-likelihood test determines whether including a random error term for the EXPER slope (ζ_{1i}) significantly improves model fit. Subsequent tests systematically involve using Eqn 10 rather than Eqn 6 and Eqn 11 rather than Eqn 7.

$$\pi_{1i} = \gamma_{10} + \gamma_{11}(\mathrm{TIB}_i) + \zeta_{1i} \quad (9)$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}(\mathrm{TIB}_i) + \zeta_{2i} \quad (10)$$

$$\pi_{3i} = \gamma_{30} + \gamma_{31}(\mathrm{TIB}_i) + \zeta_{3i} \quad (11)$$

Table 4 provides degrees of freedom, model fit indices, log-likelihood values, log-likelihood ratios and *P*-values for the model contrasts. Notice that the first two models fit the data significantly better than their predecessor based upon the −2 log-likelihood ratio test. Consequently, the best fitting model allowed for individual variability in experimental phase RT slopes and individual variability in RT change at the experimental to recovery phase transition. Individual differences in recovery slope were not present, but the power to detect such effects may be diminished by the relatively few repeated measures (three) in the recovery phase.

Step 3

In the final step of modeling the experimental design, a re-examination of the within-individual error structure was conducted because changes in the random terms can impact the nature of the within-error structure. Tests revealed evidence of lag 1 autocorrelation (−2 log-likelihood ratio = 61.76, $P < 0.0001$), but no other significant effects. Thus, a term for autocorrelation continued to be included as it had in the previous models.

Sleep restriction sample

The preceding analyses demonstrate that individual differences exist in the experimental slope and the transition phase even after design effects are controlled. To provide a more rigorous examination of differences in individual responses to sleep restriction (both in terms of overall performance and trend patterns) the results were replicated using the 3-h TIB and 5-h TIB groups as these clearly represent sleep restriction samples. The ICC based on the conditional null model (adjusting for TST) was 0.77. In addition, the log-likelihood tests (not shown) revealed reliable differences in the experimental slope and the transition parameter. These results provide evidence of significant individual differences in response to sleep restriction.

Phase 2 age analysis

Given the evidence of reliable variability in the experimental slope and transition, a series of analyses involving age were conducted on the entire sample. In the first analysis, age was included as a predictor of the experimental slope. This test was significant ($t = -3.38$, $P < 0.01$). Next age was included as a predictor of the transition. This test was also significant

Table 5 Phase 2 age model

	Parameter	*SE*	*d.f.*	*t*-value	*P*-value
Fixed effects					
Intercept (value in ms)	275.14	21.73	643	12.66	0.00
Experimental slope (EXPER)	22.48	2.15	643	10.45	0.00
Transition to recovery (TRANS)	−77.22	8.88	643	−8.69	0.00
Recovery slope (RECOV)	−13.03	3.76	643	−3.47	0.00
Average total nightly sleep time (TST)	−4.36	2.26	62	−1.93	0.06
Age	0.27	0.46	62	0.58	0.57
EXPER × TST	−2.41	0.28	643	−8.61	0.00
TRANS × TST	11.23	1.56	643	7.19	0.00
RECOV × TST	1.74	0.66	643	2.64	0.01
EXPER × Age	−0.13	0.04	643	−3.38	0.00
Variance components		Correlations			
		Intr	Exper		
Intercept (ζ)	1090.84				
Experimental slope (ζ)	12.47	0.13			
Transition to recovery (ζ)	359.79	−0.01	−0.65		
Residual (ε)	217.63				
Fit indices					
Deviance (−2 Log-likelihood)	6059.90				
AIC	6095.90				

($t = -3.15$, $P < 0.01$). Finally, age was included as a simultaneous predictor of both the experimental linear slope and the transition parameter; however, in this model the effect of age was non-significant for both parameters. This latter finding suggests shared variance between experimental slope and the transition parameter. This finding is not surprising as individuals with the steepest slopes also have the most potential for large transitional recoveries; consequently, if young participants tend to have steeper slopes they will also tend to have larger transitional changes. Nonetheless, it necessitated including only one interaction term involving age in the final model. The decision was made to interpret the interaction between age and the experimental slope because this effect causally precedes the effect associated with the transition parameter. Table 5 provides the final parameter estimates. Notice in the correlation estimates associated with the variance components that the experimental slope is negatively related to the transition parameter (−0.65) confirming that individuals with positive slopes during the experiment tend to have higher drops in RT during the transition.

Fig. 2 shows the form of the interaction involving age and experimental slope for a younger participant 29 years of age (one SD below the mean), an older participant 48 years of age (one SD above the mean) and an average participate 39 years of age in the condition where the participants slept 2.96 h per night. Both younger and older participants displayed nearly identical RTs at baseline; however, over the course of the experiment older participants showed less decline in performance than younger participants. The effect size associated with this interaction is small. Without age as a predictor, the variance estimate for the experimental slope is 12.89. With age included, the variance estimate drops to 12.47 (see Table 5). This represents a 3.2% change in explained variance. Note that effect size estimates in mixed-effects models are complicated by having several different sources of variability, so the estimate should be interpreted with caution (Singer and Willett, 2003). More details regarding the interpretation of the slope estimate will be addressed in the Discussion section.

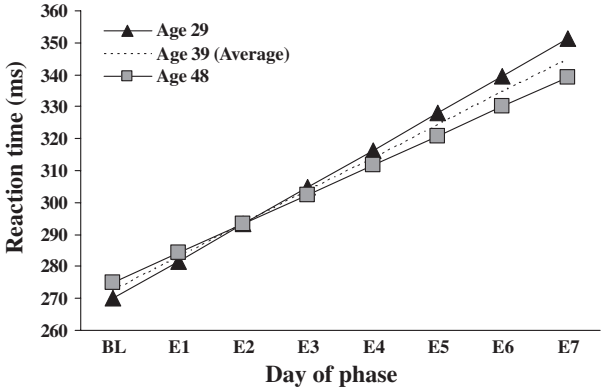

Figure 2. Psychomotor vigilance test reaction time across baseline and experimental days for a younger, average and older participant in the 3-h time in bed sleep condition.

A series of follow-on analyses examined residual variance in the experimental slope and transition parameter with both sleep hours and age as predictors to determine whether significant variability among individuals still remained. Log-likelihood analyses of nested models indicated that there was significant variability in the experimental slope and the transition parameter. These findings suggest that additional factors accounting for individual differences in performance during sleep loss remain to be discovered and modeled. Analyses were also conducted to determine whether gender effects were confounded with age effects. Recall that males tended to be older. A model adding a main effect for gender and interactions between: (a) age and gender, and (b) gender and the experimental slope failed to find any significant gender effects. Specifically, the *t*-value for the main effect for

gender was −1.80 ($P = 0.08$); the *t*-value for the age and gender interaction was 1.13 ($P = 0.26$), and the *t*-value for the gender and experimental slope interaction was 0.72 ($P = 0.47$). Importantly, however, the age effect associated with the experimental slope remained significant with the gender main effects and interactions in the model ($t = -3.07$, $P = 0.002$).

Age and vulnerability to sleep restriction

The preceding analyses demonstrate that age is related to the experimental slope such that older subjects have less of a performance decline than younger subjects. The analyses, however, do not explicitly test whether age reduces vulnerability to sleep loss. To demonstrate a protective factor of age, it is necessary to show that age interacts with the experimental slope under conditions of sleep restriction, but is unrelated to the experimental slope when subjects have no sleep restrictions. To test this proposition, a three-way interaction term involving the experimental slope (EXPER), TST and Age was added to the model presented in Table 5. The three-way interaction was non-significant ($t = 1.46$, $P = 0.15$) suggesting that age had a similar relationship to the experimental slope regardless of whether the subject was or was not sleep restricted.

DISCUSSION

Reaction time performance vulnerability associated with sleep restriction reliably varies among individuals. One way to demonstrate this effect is to show that individual performance tends, on average, to be consistent across multiple sleep deprivation sessions (Van Dongen *et al.*, 2004a). Another approach is to examine intra-individual profiles of performance over multiple periods (days) within a single long-duration session (see also Van Dongen *et al.*, 2003). In this paper, we examined intra-individual profiles of performance within a long-duration session using the framework of a discontinuous growth model. The approach allowed us to: (a) examine performance trajectories over time associated with the experimental slope, the transition period, and the recovery, and (b) identify individual factors that predict systematic differences in individual performance trajectories. This approach allowed us to further elucidate the relationship between age and vulnerability to sleep loss.

With respect to age, the results confirm that older age serves to ameliorate performance declines over repeated trials. These results are similar to the findings reported by Bonnet (1989), Bonnet and Arand (1989), Philip *et al.* (1999, 2004) and Stenuit and Kerkhofs (2005) although it is important to note that the effects of sleep loss on other more cognitively demanding measures have been found to be exacerbated by age (e.g. Webb, 1985). Unlike other studies, however, this study explicitly examined the role of age as it related to patterns of performance over time. Older individuals had less of a decline in performance during the experimental phase; however, the age effect did not significantly vary as a function of TST suggesting that age was associated with less of a performance decline regardless of whether the subject was or was not sleep restricted. Specifically, the three-way interaction among age, experimental slope and TST was non-significant. There was also evidence to suggest that younger individuals had a more pronounced recovery in the transition from experimental phase to recovery phase. This effect, although, was not completely independent of the fact that younger individuals had larger RT increases (performance declines) during the experimental phase and thereby had more opportunity to recover. Thus, the effect was not significant when considered simultaneously with the experimental phase effects.

In interpreting the age results, it has been suggested previously that age-related differences in motivation may contribute to observed differences in performance during sleep loss (Webb and Levy, 1982). While speculative, we note that the findings of the current study are congruent with a motivational explanation. The results suggest that older subjects did a better job of maintaining and/or improving performance than did their younger counterparts regardless of the amount of sleep they obtained during the experimental phase. One parsimonious explanation of this finding is older subjects may have expended more effort across days thus producing age differences in performance trends independent of TST.

In interpreting the design effects, it is important to point out that the results are congruent with results reported by Van Dongen *et al.* (2003). Specifically, Van Dongen *et al.* reported that the effects of repeated days of sleep restriction on PVT lapses were cumulative such that over time larger and larger differences emerged in the sleep-restricted relative to the non-sleep-restricted conditions. In the current study, the form of these findings were replicated such that by day 7 there were large cumulative performance decrements for those sleeping approximately 3 h per night versus those sleeping approximately 5 or 7 h (see Fig. 1). Note that the methods used in the current study are not a direct replication of Van Dongen *et al.* (2003): Van Dongen *et al.* used nonlinear approaches to identify the best fitting cumulative decline trend. Their approach allowed for greater and greater declines over time such that the incremental decline from day 5 to day 6 might be greater than the incremental decline from day 1 to day 2. Nonetheless, in both studies the gap between the sleep-restricted and non-sleep-restricted samples increased over time and these findings are congruent with the concept of cumulative decrements.

Overall, these findings serve to clarify the nature of individual differences in vulnerability to sleep loss: (a) there are reliable individual differences in overall performance and performance trends both for the sample as a whole and for the sample containing only sleep-restricted subjects, (b) a significant portion of this variability can be explained by age, but (c) even with age and experimental design effects in the model, variability remains in both the experimental slope and in the transition parameter. The discontinuous growth model provides a mechanism to empirically examine individual

differences in a number of different facets of long-duration studies (e.g. experimental phase performance, transition adaptation, and recovery performance). Note that in the current study the individual differences surrounding the transition remain unexplained. That is, even though slope and transition parameters are related, it is reasonable to expect that there may be individual difference predictors of the transition parameter that help explain variation in the transition from the experimental to recovery phase. In addition, it is reasonable to assume that there will be reliable individual differences in the recovery phase particularly if power to detect such differences is increased by designing studies with four or more days of recovery.

In absolute terms, the estimate of experimental slope variability that remaining after modeling design effects and age is relatively small compared with estimates from other parameters (see Table 5). By themselves, however, variance components have little absolute meaning (Singer and Willett, 2003). The fact that the residual slope variability estimate of 12.47 is significant, however, allows one to obtain empirical Bayes estimate predictions of the increase in RT over the experimental phase for each individual in the study (Pinheiro and Bates, 2000; Singer and Willett, 2003).

Empirical Bayes estimates use two pieces of information to calculate subject-specific predictions. First, the estimates combine information about subject-specific random effect residuals (ζ) to calculate unique parameter estimates for each individual. For instance, if the overall parameter estimate for the slope is 22.48, but the random effect residual (ζ) is -2.48 for subject 1, then subject 1's unique estimate is 20.00. Second, empirical Bayes estimates use subject-specific fixed information such as age and TST to provide individualized predictions (Singer and Willett, 2003). For instance, because the experimental slope estimate interacts with age, a subject-specific estimate for the experimental slope must also include subject age.

An examination of the subject-specific coefficients associated with the experimental slope for the 18 individuals in the 3-h TIB group revealed a range of 15.04-ms increase per day to a 28.89-ms increase per day. Using this information, and assuming fixed effects with: (a) constant 3-h TST, and (b) subject-specific age revealed that in a 7-day study the least vulnerable individual had an 11.23-ms increase while the most vulnerable individual had a 102.07-ms increase. Part of this difference reflects differences in age (the least vulnerable individual was 46 and the most vulnerable individual was 36), but age effects are relatively weak as evident in Fig. 2. Consequently, the 90.84-ms cumulative difference largely reflects unknown individual differences. In interpreting this difference recall that the 90.84-ms difference roughly equates to the large difference induced by the design effect presented in Fig. 1 for 2.96 TST versus 6.78 TST. Empirical Bayes estimates for the 5-h TIB group predicted a 23.56-ms increase in RT over the 7 days for the least vulnerable individual (31 years old), and a 109.79-ms increase for the most vulnerable (also a 31 year old). Thus, it is clear that the variance estimate of 12.47 in Table 5 represents considerable individual variability in susceptibility to sleep restriction over the course of a 7-day study.

Extensions and conclusions

The current study illustrated how the discontinuous growth model, a variant of the mixed-effects model, could be used to examine individual differences in performance trends within a long-duration sleep study. Variants of mixed-effects models hold considerable promise for studying individual differences in response to sleep restriction as illustrated by this study and by the work of Van Dongen and colleagues.

We note that the flexibility of mixed-effects models allows them to be extended in a variety of ways. One important extension is the ability to include time-varying covariates. Specifically, in mixed-effects models, it is possible to predict repeated measure outcomes using covariates that also repeatedly vary (Singer and Willett, 2003). For instance, in the current example, daily RT could have been predicted using sleep hours from the previous night in addition to or in lieu of the three variants of time (experimental slope, transition, and recovery slope). If the objective of a study is to identify the strongest predictors of variability among repeated measures, researchers should consider including time-varying covariates. In conclusion, we anticipate that applying mixed-effects models to sleep studies will continue to provide considerable information about the effects of sleep restriction and the nature of individual differences.

ACKNOWLEDGEMENTS

This work was supported by the US Army Medical Research and Material Command (Military Operational Medicine Program, Project S15 Q), the US Department of Transportation (Contract/Grant No. DTFH61-94-Y-00090), and the GCRC/Johns Hopkins Bayview Medical Center (Grant No. M01RR02719), Baltimore, MD. This material has been reviewed by the Walter Reed Army Institute of Research, and there is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the position of the Department of the Army or the Department of Defense. This study was approved by the Walter Reed Army Institute of Research Human Use Committee and the US Army Medical Research and Materiel Command Human Subjects Review Board of the Army Surgeon General and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

REFERENCES

Balkin, T., Thorne, D., Sing, H., Thomas, M., Redmond, D., Wesensten, N., Williams, J., Hall, S. and Belenky, G. *Effects of Sleep Schedules on Commercial Driver Performance*. Report No. DOT-MC-00–133. US Department of Transportation, Federal Motor Carrier Safety Administration, Washington, DC, 2000.

Balkin, T. J., Bliese, P. D., Belenky, G., Sing, H., Thorne, D. R., Thomas, M., Redmond, D. P., Russo, M. and Wesensten, N. J. Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *J. Sleep Res.*, 2004, 13: 219–227.

Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., Russo, M. B. and Balkin, T. J. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose–response study. *J. Sleep Res.*, 2003, 12: 1–13.

Bliese, P. D. and Ployhart, R. E. Growth modeling using random coefficient models: model building, testing, and illustrations. *Org. Res. Methods*, 2002, 5: 362–388.

Bonnet, M. H. The effect of sleep fragmentation on sleep and performance in younger and older subjects. *Neurobiol. Aging*, 1989, 10: 21–25.

Bonnet, M. H. and Arand, D. L. Sleep loss in aging. *Clin. Geriatr. Med.*, 1989, 5: 405–420.

Bonnet, M. H. and Rosa, R. R. Sleep and performance in young adults and older insomniacs and normals during acute sleep loss and recovery. *Biol. Psychol.*, 1987, 25: 153–172.

Hox, J. J. *Multilevel Analysis: Techniques and Applications*. Larwence Erlbaum Associates, Mahwah, NJ, 2002.

Littell, R. C., Milliken, G. A., Stroup, W. W. and Wolfinger, R. D. *SAS System for Mixed Models*. SAS Institute Inc., Cary, NC, 1996.

Olofsen, E., Dinges, D. F. and Van Dongen, H. P. A. Non-linear mixed effects modeling: individualization and prediction. *Aviat. Space Environ. Med.*, 2004, 75: 134–140.

Philip, P., Taillard, J., Quera-Salva, M. A., Bioulac, B. and Akerstedt, T. Simple reaction time, duration of driving and sleep deprivation in young versus old automobile drivers. *J. Sleep Res.*, 1999, 8: 9–14.

Philip, P., Taillard, J., Sagaspe, P., Valtat, C., Sanchez-Ortuno, M., Moore, N., Charles, A. and Bioulac, B. Age, performance and sleep deprivation. *J. Sleep Res.*, 2004, 13: 105–110.

Pinheiro, J. C. and Bates, D. M. *Mixed-effects Models in S and S-PLUS*. Springer-Verlag, New York, 2000.

Raudenbush, S. W. and Bryk, A. S. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage, Newbury Park, CA, 2002.

R Development Core Team. *R: A Language and Environment for Statistical Computing, Reference Index Version 2.2.0*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available at: http://www.R-project.org., 2005.

Rechtschaffen, A. and Kales, A. *A Manual of Standardized Techniques and Scoring System for Sleep Stages of Human Subjects*. Brain Information Service and Brain Research Institute, Los Angeles, CA, 1968.

Singer, J. D. and Willett, J. B. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, New York, 2003.

Stenuit, P. and Kerkhofs, M. Age modulates the effects of sleep restriction in women. *Sleep*, 2005, 28: 1283–1288.

Van Dongen, H. P. A., Maislin, G., Mullington, J. M. and Dinges, D. F. The cumulative cost of additional wakefulness: dose–response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 2003, 26: 117–126.

Van Dongen, H. P. A., Baynard, M. D., Maislin, G. and Dinges, D. F. (a) Systematic inter-individual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep*, 2004a, 27: 423–433.

Van Dongen, H. P. A., Maislin, G. and Dinges, D. F. (b). Dealing with inter-individual differences in the temporal dynamics of fatigue and performance: importance and techniques. *Aviat. Space Environ. Med.*, 2004b, 75: 147–154.

Webb, W. B. A further analysis of age and sleep deprivation effects. *Psychophysiology*, 1985, 22: 156–161.

Webb, W. B. and Levy, C. M. Age, sleep deprivation, and performance. *Psychophysiology*, 1982, 19: 272–276.