



## Original Contribution

### Long Working Hours and Cognitive Function

#### The Whitehall II Study

**Marianna Virtanen, Archana Singh-Manoux, Jane E. Ferrie, David Gimeno, Michael G. Marmot, Marko Elovainio, Markus Jokela, Jussi Vahtera, and Mika Kivimäki**

*Initially submitted June 5, 2008; accepted for publication November 3, 2008.*

This study examined the association between long working hours and cognitive function in middle age. Data were collected in 1997–1999 (baseline) and 2002–2004 (follow-up) from a prospective study of 2,214 British civil servants who were in full-time employment at baseline and had data on cognitive tests and covariates. A battery of cognitive tests (short-term memory, Alice Heim 4-I, Mill Hill vocabulary, phonemic fluency, and semantic fluency) were measured at baseline and at follow-up. Compared with working 40 hours per week at most, working more than 55 hours per week was associated with lower scores in the vocabulary test at both baseline and follow-up. Long working hours also predicted decline in performance on the reasoning test (Alice Heim 4-I). Similar results were obtained by using working hours as a continuous variable; the associations between working hours and cognitive function were robust to adjustments for several potential confounding factors including age, sex, marital status, education, occupation, income, physical diseases, psychosocial factors, sleep disturbances, and health risk behaviors. This study shows that long working hours may have a negative effect on cognitive performance in middle age.

cognition; middle aged; prospective studies; vocabulary; work

Abbreviations: AH 4-I, Alice Heim 4-I; GHQ-30, 30-item General Health Questionnaire.

Long working hours are common worldwide; for example, in the European Union member states, 12%–17% of employees worked overtime in 2001 (1). Long working hours have been found to be associated with cardiovascular and immunologic reactions, reduced sleep duration, unhealthy lifestyle (2–8), and adverse health outcomes, such as cardiovascular disease, diabetes, subjective health complaints, fatigue (2–7), and depression (8). There is increasing evidence to suggest the importance of midlife risk factors for later dementia (9). Furthermore, the link between cognitive impairment and later life dementia is clearly established (10, 11). Thus, it is important to examine risk factors for poor cognition in midlife, and there is little research on the potential effects of long working hours on cognition among middle-aged persons.

A cross-sectional study of 248 automotive workers found an association between overtime work and impaired perfor-

mance on tests of attention and executive function (12). This finding was in agreement with findings from other studies that focused on different forms of shift work or work schedule rather than on long working hours (13, 14). For example, deterioration in cognitive performance, including impaired grammatical reasoning and alertness, has been found in post versus pretest conditions among employees working 9- to 12-hour shifts compared with a traditional 8-hour shift (13). However, little is known about the health effects of long total working hours as opposed to long hours of shift work.

This study examined the relation between long working hours and cognitive function over a 5-year follow-up period in a large-scale, prospective occupational cohort of British civil servants (the Whitehall II study) (15). We were able to take into account several factors that may act as confounders or mediators of this association, such as education, occupational position, physical health status, psychological

and psychosocial factors, sleep problems, and health risk behaviors (2).

## MATERIALS AND METHODS

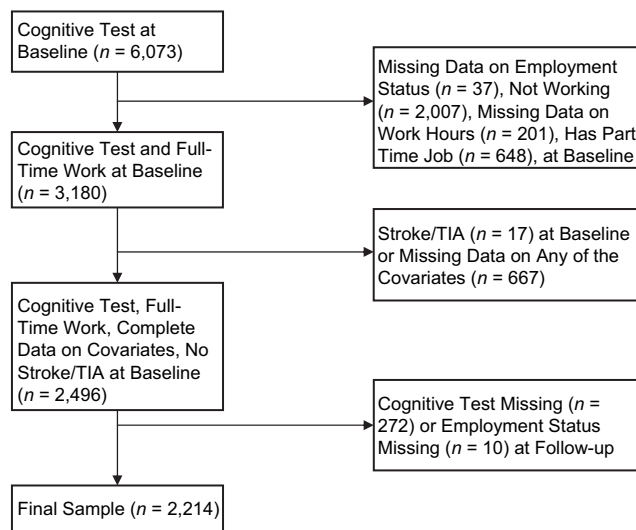
### Participants and procedure

The Whitehall II study sample recruitment (phase 1) took place between late 1985 and early 1988 among all office staff, aged 35–55 years, from 20 London-based Civil Service departments (15). The response rate was 73% (6,895 men and 3,413 women). Since phase 1, there have been 7 further data collection phases. Informed consent was gained from all participants. The University College London Medical School Committee on the Ethics of Human Research approved the protocol.

As cognitive performance was measured on the whole sample for the first time at phase 5, this phase is used as baseline for the present study. We included all 2,214 participants (1,694 men and 520 women) who were employed and responded to the questions on working hours and for whom the covariates and cognitive test scores were available at phase 5 (1997–1999) and phase 7 (2002–2004). A flow chart of sample selection is shown in Figure 1. The mean age of the 2,214 participants at phase 5 was 52.1 years (standard deviation, 4.2; range, 45–66). There were no major differences between the participants and all full-time employees who participated in phase 5 ( $n = 3,597$ ) in terms of age (52.1 vs. 52.4 years), sex (77% vs. 75% male), occupational grade (18% with the lowest occupational grade vs. 22%), and prevalence of coronary heart disease (10% vs. 11%). However, employees who participated in our study at phases 5 and 7 differed from the cohort at recruitment to the Whitehall II study ( $n = 10,308$ ), in that they were younger (mean age, 40.6 vs. 44.5 years at phase 1); more likely to be male (77% vs. 67%) and from the higher socioeconomic groups (10% with the lowest grade vs. 23%); and less likely to have pre-existing coronary heart disease at phase 1 (2.7% vs. 4.1%).

### Tests of cognitive function

The cognitive function test battery at phases 5 and 7 consisted of 5 standard tasks chosen to evaluate cognitive functioning in middle-aged adults. The first was verbal memory assessed by a 20-word free recall test of short-term memory. Participants were presented a list of 20 one- or two-syllable words at 2-second intervals and were then asked to recall in writing as many of the words in any order within 2 minutes. The Alice Heim 4-I (AH 4-I) test (16) is a measure of inductive reasoning that assesses fluid intelligence, that is, the ability to identify patterns and to infer principles and rules. This test is composed of a series of 65 items (32 verbal and 33 mathematical reasoning items) of increasing difficulty. The participants had 10 minutes to complete this section. The Mill Hill vocabulary test (17) assesses crystallized intelligence, that is, knowledge of verbal meaning, and encompasses the ability to recognize and comprehend words. We used this test in its multiple-choice format that consists of a list of 33 stimulus words ordered by increasing difficulty, with 6 response choices per word. The



**Figure 1.** Sample selection, the Whitehall II study, 1997–2004. TIA, transient ischemic attack.

final 2 tests were measures of verbal fluency: phonemic and semantic (18). Phonemic fluency was assessed via “S” words, and semantic fluency was assessed via “animal” words. Subjects were asked to recall in writing as many words beginning with “S” and as many animal names as they could. One minute was allowed for each test of verbal fluency. A higher score indicated better performance in each test.

The change score was calculated for each measure of cognitive function as phase 7 score minus phase 5 score. As the time interval between clinical examination at phases 5 and 7 varied between 3.9 and 7.1 years (mean, 5.5 years), the difference in cognitive score was divided by the time in years between the 2 measures for each individual and multiplied by 5 to give everyone the same (5-year) time period between the 2 phases of cognitive data collection.

### Working hours and other baseline characteristics

Working hours were determined at phase 5 from the following 2 questions: “How many hours do you work per average week in your main job including work brought home?” and “How many hours do you work in an average week in your additional employment?”. Participants were divided into the following 3 groups: a total of 35–40 hours; 41–55 hours; and more than 55 hours per week (5–7). In addition, analyses were conducted by using the scale as a continuous variable. Participants in the Whitehall II study are almost exclusively white-collar civil servants. The most common weekly working hours correspond to 36 hours per week net, although various flexible working arrangements can also be arranged. In the present cohort, the total mean working hours were 45.2 hours/week (standard deviation, 8.0; range, 35–120).

Altogether, 20 sociodemographic characteristics and behavioral, psychological, psychosocial, and medical conditions

known to be associated with cognitive function and/or working hours were included as covariates in the analysis (2–9, 12, 19–38). In addition to sex and age, marital status, indicators of socioeconomic position, that is, occupational grade (6 levels from which the lowest 2 levels were collapsed to obtain sufficient numbers), education (postgraduate, graduate, higher secondary school, lower secondary school, or no academic qualifications), and the participant's report of his/her annual gross salary were assessed. Employment status (working vs. not working) at follow-up was obtained from the phase 7 questionnaire.

The physical functioning component score of the Medical Outcomes Study SF-36 test (39) was used as a measure of global physical health status and divided into quartiles separately for men and women. Prevalent coronary heart disease at phase 5 included cases of nonfatal myocardial infarction and angina. In addition to definite nonfatal myocardial infarction and definite angina, our total nonfatal coronary heart disease events outcome included self-reported cases in the absence of any clinical record evidence of coronary disease. Systolic blood pressure and diastolic blood pressure were measured by using a Hawksley random-zero sphygmomanometer (Hawksley and Sons, Ltd., Lancing, United Kingdom). In keeping with standard definitions, subjects with systolic blood pressure of  $\geq 140$  mm Hg and diastolic blood pressure of  $\geq 90$  mm Hg or on antihypertensive treatment were considered to be hypertensive (40).

Psychological distress was assessed by using the 30-item General Health Questionnaire (GHQ-30) (41). The GHQ-30 has been validated in a number of diverse populations and has been validated specifically against the Clinical Interview Schedule in Whitehall II data, giving a cutoff point of 4/5 positive responses for dividing noncases from cases (42). In addition, a 5-item subscale of anxiety (e.g., feelings of constant strain, panic, nervousness) was derived from the GHQ-30 (41). Scores in the top decile were used to define anxiety cases, corresponding to the prevalence of anxiety disorders in the general population (43).

Sleep was assessed in 2 ways; the first was a measure of duration with respondents identified as short sleepers if they reported sleeping less than 6 hours on an average week night (44). Sleep quality was assessed by using the "Jenkins scale" (45), which assesses sleep disturbances during the past 4 weeks. The mean response score for all 4 questions was divided into tertiles.

Of the health behaviors, alcohol consumption (units/week) was classified into 3 categories: none;  $>0$ –14 (women)/21 (men) units;  $>14$ / $>21$  units (46). Smoking was assessed by a single question of whether the respondent was a current smoker or not. For the physical activity score, the participants were asked about the frequency and duration of their participation in physical activity (47). The amount of time spent in activities with metabolic equivalent values ranging from 0 to 6 or above was summed to allow calculation of the total number of hours per week of physical activity and divided into 3 categories—low, moderate, and high.

Social support was measured by the 15-item Close Persons Questionnaire (48), which includes questions about confiding/emotional support, practical support, and negative aspects of close relationships. The mean of all responses

was divided into tertiles. Strain in family relations was measured with a single-item question of how often the participant had any worries or problems with other relatives, for example, parents or in-laws (always/often vs. sometimes/seldom/never/not applicable). Job strain was formulated by splitting the job demands score and decision latitude score at their medians. High demands and low decision latitude indicated high job strain, and other combinations indicated low job strain (49).

### Statistical analysis

All analyses were carried out by using SAS, version 9.1, statistical software (SAS Institute, Inc., Cary, North Carolina), except missing-data analysis which was done using STATA, version 9.0, statistical software (StataCorp LP, College Station, Texas). First, we compared baseline characteristics of the participants by working hours and compared the longer-hours group ( $>55$  hours per week) with the employees with normal working hours (35–40 hours per week) using  $\chi^2$  tests. We used multiple analysis of covariance to examine whether work hours had an overall association with cognitive function, as checking for each measure of cognitive function separately increases the chance of Type I error. Subsequently, analysis of variance was used to assess the association between work hours and individual measures of cognitive function. When a significant difference was found in cognitive function tests at baseline and/or at follow-up between groups, additional analyses were carried out with the change score to assess temporal order and to examine whether the change was statistically significant. Sequential analyses were undertaken to see whether adjustment for covariates attenuated the association between long working hours and change in cognitive function. Age was entered into the models as a continuous variable, and all other covariates were entered as categorical variables. As recommended by Glymour et al. (50), we used baseline-unadjusted change scores for cognitive change. In order to examine linear trend in the association between working hours and cognitive function, we repeated the analysis using working hours as a continuous variable.

To explore whether selection bias might have occurred because of loss to follow-up, we undertook a sensitivity analysis in which we used multiple multivariate imputation (51) using working hours, all covariates, and cognition variables to impute values for missing values in any variables with some missing data, among all 3,163 participants free of stroke and transient ischemic attack at baseline. We used switching regression in STATA software, as described by Royston (51), carried out 20 cycles of regression switching, and generated 20 imputation data sets. The multiple multivariate imputation approach creates a number of copies of the data (in this case, we generated 20 copies), each of which has values that are missing imputed with an appropriate level of randomness using chained equations. The estimates are obtained by averaging across the results from each of these 20 data sets using Rubin's rules. The procedure takes account of uncertainty in the imputation, as well as uncertainty due to random variation, as undertaken in all multivariable analyses.

**Table 1.** Characteristics of the Participants by Working Hours at Baseline, the Whitehall II Study, 1997–2004

	All		Working Hours per Week						P Value <sup>a</sup>
			≤40		41–55		>55		
	No.	%	No.	%	No.	%	No.	%	
Sex									
Men	1,694	77	607	71	936	79	151	83	<0.001
Women	520	23	246	29	244	21	30	17	
Age, mean years (SE)	52.1 (0.09)		52.4 (0.14)		51.8 (0.12)		52.5 (0.31)		0.741
Marital status									
Married/cohabited	1,749	79	624	73	969	82	156	86	<0.001
Nonmarried/noncohabited	465	21	229	27	211	18	25	14	
Occupational grade level									
1 (highest)	495	22	75	9	328	28	92	51	<0.001
2	569	26	164	19	356	30	49	27	
3	359	16	161	19	180	15	18	10	
4	400	18	231	27	160	14	9	5	
5–6 (lowest)	391	18	222	26	156	13	13	7	
Educational level									
Postgraduate	378	17	93	11	238	20	47	26	<0.001
Graduate	551	25	200	23	294	25	57	31	
Higher secondary	657	30	251	29	356	30	50	28	
Lower secondary	499	23	240	28	235	20	24	13	
No academic qualifications	129	6	69	8	57	5	3	2	
Income, £/year									
≥50,000	358	16	46	5	232	20	80	44	<0.001
25,000–<50,000	1,176	53	393	46	703	60	80	44	
15,000–<25,000	579	26	351	41	214	18	14	8	
<15,000	101	5	63	7	31	3	7	4	
Physical health status									
I (lowest)	404	18	148	17	215	18	41	23	0.137
II	551	25	219	26	298	25	34	19	
III	589	27	234	27	301	26	54	30	
IV (highest)	670	30	252	30	366	31	52	29	
Coronary heart disease									
No	1,988	90	769	90	1,059	90	160	88	0.478
Yes	226	10	84	10	121	10	21	12	
Hypertension									
No	1,801	81	682	80	969	82	150	83	0.368
Yes	413	19	171	20	211	18	31	17	
Psychological distress									
No	1,674	76	678	79	880	75	116	64	<0.001
Yes	540	24	175	21	300	25	65	36	

Table continues

## RESULTS

Characteristics of the study participants by working hours at baseline are shown in Table 1. A total of 853 (39%) participants reported 35–40 hours of work per week, 1,180 (53%) reported 41–55 hours, and 181 (8%) reported more

than 55 hours of work per week. Compared with employees with 35–40 hours, a higher percentage of those who worked more than 55 hours were men and were married or cohabited and had a higher occupational grade, higher education, higher income, more psychological distress, shorter sleep, higher alcohol use, and more social support.

Table 1. Continued

	All		Working Hours per Week						P Value <sup>a</sup>
			≤40		41–55		>55		
	No.	%	No.	%	No.	%	No.	%	
Anxiety									
No	1,972	89	770	90	1,044	88	158	87	0.230
Yes	242	11	83	10	136	12	23	13	
Short sleep (<6 hours)									
No	2,043	92	799	94	1,084	92	160	88	0.013
Yes	171	8	54	6	96	8	21	12	
Sleeping problems									
Low	707	32	292	34	360	31	55	30	0.375
Intermediate	836	38	316	37	455	39	65	36	
High	671	30	245	29	365	31	61	34	
Alcohol use									
No	259	12	118	14	127	11	14	8	0.018
Moderate	1,382	62	553	65	714	61	115	64	
High	573	26	182	21	339	29	52	29	
Smoking									
No	2,020	91	777	91	1,076	91	167	92	0.611
Yes	194	9	76	9	104	9	14	8	
Physical activity									
Low	352	16	153	18	164	14	35	19	0.819
Intermediate	780	35	326	38	389	33	65	36	
High	1,082	49	374	44	627	53	81	45	
Social support									
Low	763	34	312	37	406	34	45	25	0.008
Intermediate	764	35	306	36	386	33	72	40	
High	687	31	235	28	388	33	64	35	
Strain in family relations									
No	1,792	81	700	82	951	81	141	78	0.192
Yes	422	19	153	18	229	19	40	22	
Job strain									
No	1,671	75	648	76	886	75	137	76	0.937
Yes	543	25	205	24	294	25	44	24	
Employment status at follow-up									
Employed	1,680	76	624	73	911	77	36	80	0.052
Nonemployed	534	24	229	27	269	23	145	20	

Abbreviation: SE, standard error.

<sup>a</sup> P value for difference between the groups working 40 hours or less and those working more than 55 hours per week.

Multiple analysis of covariance revealed an overall association of working hours with cognitive function at baseline ( $P = 0.002$ ) and follow-up ( $P = 0.037$ ), as well as change in cognitive function scores between baseline and follow-up ( $P = 0.044$ ). Table 2 shows the associations between working hours at baseline and each cognitive function measure at baseline and at follow-up after adjustment for all the covariates measured at baseline. Compared with employees working 40 hours or less per week, employees working

more than 55 hours had lower vocabulary scores at baseline and at follow-up. At follow-up, they had lower scores also on the reasoning test. No significant difference between groups was found in any other measures of cognitive function at follow-up. Repeating these analyses with working hours treated as a continuous variable largely replicated the findings and additionally showed an association between working hours and better phonemic fluency at baseline but not at follow-up.

**Table 2.** Association Between Working Hours at Baseline and Cognitive Function at Baseline and at Follow-up, Fully Adjusted Models,<sup>a</sup> the Whitehall II Study, 1997–2004<sup>b</sup>

Weekly Working Hours at Baseline	Memory Range, 0–18 <sup>c</sup> (1–18) <sup>d</sup>		Reasoning Range, 12–65 <sup>c</sup> (10–65) <sup>d</sup>		Vocabulary Range, 1–33 <sup>c</sup> (6–32) <sup>d</sup>		Phonemic Fluency Range, 3–47 <sup>c</sup> (2–34) <sup>d</sup>		Semantic Fluency Range, 2–34 <sup>c</sup> (2–33) <sup>d</sup>	
	Mean (SE)	<i>P</i> Value <sup>e</sup>	Mean (SE)	<i>P</i> Value <sup>e</sup>	Mean (SE)	<i>P</i> Value <sup>e</sup>	Mean (SE)	<i>P</i> Value <sup>e</sup>	Mean (SE)	<i>P</i> Value <sup>e</sup>
<i>Cognitive function at baseline</i>										
≤40	6.94 (0.18)	Referent	46.14 (0.63)	Referent	24.80 (0.25)	Referent	16.95 (0.33)	Referent	16.83 (0.30)	Referent
41–55	7.12 (0.17)	0.081	46.02 (0.60)	0.744	24.38 (0.24)	0.005	17.25 (0.31)	0.117	16.87 (0.29)	0.810
>55	7.14 (0.23)	0.306	45.93 (0.79)	0.763	23.96 (0.32)	0.002	17.62 (0.41)	0.056	17.08 (0.38)	0.441
Test for linear trend <sup>f</sup>	<i>P</i> = 0.835		<i>P</i> = 0.206		<i>P</i> < 0.001		<i>P</i> = 0.031		<i>P</i> = 0.874	
<i>Cognitive function at follow-up</i>										
≤40	7.11 (0.18)	Referent	44.17 (0.65)	Referent	24.97 (0.25)	Referent	15.66 (0.31)	Referent	16.20 (0.28)	Referent
41–55	7.18 (0.18)	0.547	43.53 (0.62)	0.099	24.62 (0.24)	0.020	15.95 (0.29)	0.111	16.18 (0.26)	0.912
>55	6.93 (0.23)	0.359	42.74 (0.81)	0.040	24.39 (0.32)	0.032	16.00 (0.38)	0.302	16.08 (0.35)	0.680
Test for linear trend <sup>f</sup>	<i>P</i> = 0.118		<i>P</i> = 0.010		<i>P</i> = 0.003		<i>P</i> = 0.088		<i>P</i> = 0.430	

Abbreviation: SE, standard error.

<sup>a</sup> Adjusted for age, sex, marital status, follow-up employment status, occupational grade, education, income, physical health indicators, psychological distress, anxiety, sleep problems, health risk behaviors, social support, family stress, and job strain.

<sup>b</sup> In each cognitive test, a higher score indicates better cognitive performance.

<sup>c</sup> Range of scores at baseline.

<sup>d</sup> Range of scores at follow-up.

<sup>e</sup> *P* value for difference with the referent group working 40 hours or less per week.

<sup>f</sup> Total working hours entered into the model as a continuous variable.

Table 3 examines the mean difference in the change in reasoning score between those working normal hours and those working long hours. Successive models show the effects of step-by-step adjustments. The stepwise adjustments show that various adjustments produced little attenuation of the effect of working hours on the decline in reasoning score, and a clear dose-response pattern was revealed between exposure and outcome. Again, the findings were replicated in models replacing categories with a continuous measure of working hours.

To further examine whether the findings are robust, we ran a sensitivity analysis in a subgroup of participants still employed at follow-up ( $n = 1,672$ ,  $n = 1,677$ ). Consistent with the main analyses, working more than 55 hours versus 40 hours or less was associated with a greater decline in the reasoning score (difference,  $-1.47$ ;  $P = 0.002$ ) and lower scores on the vocabulary test at baseline (difference,  $-0.77$ ;  $P = 0.009$ ) and at follow-up (difference,  $-0.60$ ;  $P = 0.046$ ). Corresponding *P* values for the continuous working hours were  $P = 0.009$ ,  $P = 0.004$ , and  $P = 0.023$ .

To examine sex differences, we conducted altogether 15 tests of interaction between sex and continuous working hours on cognitive function outcomes and found 2 statistically significant interactions: for the vocabulary test at baseline ( $P = 0.015$ ) and at follow-up ( $P = 0.003$ ). Sex-stratified analysis showed a significant negative association between working hours and vocabulary score at baseline and at follow-up among men ( $P < 0.001$ ) but not among women ( $P = 0.899$  and  $0.339$ ).

Finally, Table 4 repeats the analyses on those associations that were found to be robust in Tables 2 and 3, except that the results were obtained from the multiple multivariate im-

putation analysis for the baseline population, a total of 3,163 participants. To simplify comparison of cohorts before and after imputations, we present the effects of working hours as per 10-hour increase in a continuous measure. Imputation had little effect on the associations with vocabulary at baseline and follow-up and with reasoning at follow-up. The association with reasoning at baseline was strengthened, but otherwise the associations were similar to those before imputation. Corresponding *P* values for the categorical working hours variable were as follows: Between the groups of >55 hours versus ≤40 hours,  $P < 0.001$  for the vocabulary score at baseline and follow-up;  $P = 0.068$  for the reasoning score at baseline;  $P = 0.002$  for the reasoning score at follow-up; and  $P = 0.025$  for the change score in reasoning (data not shown), thus replicating the original findings.

## DISCUSSION

In this study of middle-aged men and women, working more than 55 hours per week was associated with lower scores on 2 of the 5 tests of cognitive function. Long working hours at baseline were related to poorer performance on the vocabulary test at both baseline and follow-up. Furthermore, long working hours predicted decline in performance on the reasoning test over a 5-year follow-up period. These effects were robust to adjustments for 20 potential confounding factors, such as education, occupational position, physical diseases (cardiovascular dysfunction), psychosocial stress factors, sleep problems, and health risk behaviors.

We found an association between long working hours and decline in the scores for the AH 4-I reasoning test and

**Table 3.** Stepwise Adjustments<sup>a</sup> of the Association Between Working Hours at Baseline and Change in Reasoning Score Between Baseline and Follow-up, the Whitehall II Study, 1997–2004

Weekly Working Hours at Baseline	Change in Reasoning Score											
	Model I		Model II		Model III		Model IV		Model V		Model VI	
	Mean Difference (SE)	P Value <sup>b</sup>	Mean Difference (SE)	P Value <sup>b</sup>	Mean Difference (SE)	P Value <sup>b</sup>	Mean Difference (SE)	P Value <sup>b</sup>	Mean Difference (SE)	P Value <sup>b</sup>	Mean Difference (SE)	P Value <sup>b</sup>
≤40	-1.90 (0.19)	Referent	-1.95 (0.24)	Referent	-1.71 (0.28)	Referent	-1.94 (0.25)	Referent	-1.87 (0.21)	Referent	-1.77 (0.39)	Referent
41–55	-2.36 (0.18)	0.033	-2.41 (0.23)	0.035	-2.18 (0.27)	0.030	-2.37 (0.24)	0.049	-2.34 (0.20)	0.031	-2.23 (0.37)	0.046
>55	-2.97 (0.38)	0.006	-2.98 (0.41)	0.009	-2.80 (0.42)	0.006	-2.99 (0.41)	0.008	-2.96 (0.39)	0.006	-2.90 (0.49)	0.007
Mean difference <sup>c</sup>	-1.08		-1.04		-1.09		-1.05		-1.09		-1.13	
Test for linear trend <sup>d</sup>		P = 0.020		P = 0.026		P = 0.018		P = 0.026		P = 0.018		P = 0.036

Abbreviation: SE, standard error.

<sup>a</sup> Model I, adjusted for age, sex, marital status, and follow-up employment status. Model II, model I adjustments plus additionally adjusted for physical health indicators. Model III, model I adjustments plus additionally adjusted for psychological distress, anxiety, and sleep problems. Model IV, model I adjustments plus additionally adjusted for health risk behaviors. Model V, model I adjustments plus additionally adjusted for social support, family stress, and job strain. Model VI, adjusted for all covariates in models I, II, III, IV, and V plus occupational grade, education, and income.

<sup>b</sup> P for difference with the referent group working 40 hours or less per week.

<sup>c</sup> Difference in mean between the group working more than 55 hours and the group working 40 hours or less per week.

<sup>d</sup> Total working hours entered into the model as a continuous variable.

associations with the Mill Hill vocabulary tests at baseline and at follow-up. The AH 4-I test is also recognized as a measure of fluid intelligence, that is, executive function or “meta” cognitive ability as it integrates other cognitive processes such as memory, attention, and speed of information processing. Fluid intelligence is seen to be intrinsically associated with information processing and involves short-term memory, abstract thinking, creativity, ability to solve novel problems, and reaction time. It is the aspect of intelligence most affected by aging, biologic factors, diseases, and injuries (52, 53). Fluid intelligence usually increases up to the mid-20s, after which it gradually declines until the 60s when a more rapid decline takes place.

The Mill Hill vocabulary test measures crystallized intelligence that is assumed to accumulate during the lifespan through education, occupational and cultural experience, and exposure to culture and intellectual pursuits (52, 53). Crystallized abilities usually increase up to the sixth or seventh decade of age and may not decrease until after 80 years of age. We found the Mill Hill scores to remain relatively stable as expected for this middle-aged cohort. However, the Mill Hill scores were lower among employees with long working hours at both baseline and follow-up. This consistency with 2 separate measures with a 5-year interval suggests not only a plausible finding but also stability of the far-reaching effect of long working hours on vocabulary. We did not find an interaction effect between follow-up employment status and working hours on significant outcomes, which suggests that the associations found are not dependent on employment status at follow-up. However, people who work long hours might be exposed to a narrower variation of intellectual pursuits, that is, only to those that are related to their work tasks, and therefore might not be able to develop a wide variety of functions in crystallized intelligence measured by the test. However, reversed causality is also possible: Employees with lower cognitive ability may be more prone to work overtime than workers with good cognitive ability in order to get their work done.

Previous literature, mostly cross-sectional, suggests that long working hours are associated with various health outcomes, the strongest effects being observed for cardiovascular diseases, fatigue, and sleep disturbances (2–8). These can also be hypothesized to be mediating mechanisms for the association between long working hours and cognitive decline. Hypertension is associated with cognitive dysfunction by producing subtle disturbances in cerebral perfusion and affecting brain cell metabolism (19, 20). However, we found no evidence of an association between long working hours and hypertension or coronary heart disease, suggesting that the effect of long hours on cardiovascular dysfunction, if any, is unlikely to explain cognitive decline in this study.

Another hypothesis on mediating mechanisms links long working hours with psychological stress and poor recovery from work as indicated by sleeping problems and reduced sleep. Psychological stress has been suggested as affecting the brain via 2 neuroendocrine systems: 1) the sympathetic adrenomedullary system with the secretion of epinephrine and norepinephrine and 2) the hypothalamic-pituitary-adrenocortical system with the secretion of cortisol (54). Of the few studies in the field, only 1 study has found an

**Table 4.** Multivariable-adjusted<sup>a</sup> Associations Among Working Hours, Vocabulary, and Reasoning for Participants Before and After Imputation of Missing Data, the Whitehall II Study, 1997–2004

Weekly Working Hours at Baseline	Vocabulary Score <sup>b</sup>				Reasoning Score <sup>c</sup>					
	Baseline		Follow-up		Baseline		Follow-up		Change	
	Beta (SE)	P Value	Beta (SE)	P Value	Beta (SE)	P Value	Beta (SE)	P Value	Beta (SE)	P Value
	<i>Before imputation</i>									
Per 10-hour increase <sup>d</sup>	−0.38 (0.09)	<0.001	−0.27 (0.09)	0.003	−0.28 (0.23)	0.206	−0.60 (0.23)	0.010	−0.30 (0.14)	0.036
	<i>After imputation<sup>e</sup></i>									
Per 10-hour increase <sup>d</sup>	−0.43 (0.08)	<0.001	−0.37 (0.09)	<0.001	−0.54 (0.20)	0.005	−0.83 (0.21)	<0.001	−0.28 (0.13)	0.033

Abbreviation: SE, standard error.

<sup>a</sup> Adjusted for age, sex, marital status, follow-up employment status, occupational grade, education, income, physical health indicators, psychological distress, anxiety, sleep problems, health risk behaviors, social support, family stress, and job strain.

<sup>b</sup> Vocabulary score:  $n = 2,210$  and  $n = 3,163$  before and after imputation, respectively.

<sup>c</sup> Reasoning score:  $n = 2,204$  and  $n = 3,163$  before and after imputation, respectively.

<sup>d</sup> Continuous variable for working hours.

<sup>e</sup> Based on multiple multivariate imputations.

association between long working hours and neuroendocrinologic stress markers (55). We found that long working hours were associated with short sleep duration and psychological distress but not with sleep disturbances. Further adjustment for these factors did not provide support for the hypothesis that psychological distress and poor recovery act as mediating mechanisms.

The third hypothesis suggests that long working hours may affect cognitive function through health risk behaviors. Evidence on the association between long working hours and unhealthy behaviors is weak, but there is stronger evidence for the relation between health behaviors and cognitive function (22–24, 26). We found that adjustment for all these health risk behaviors had no effect on the association between long working hours and cognitive function, suggesting that health risk behaviors may not be an important mediating or confounding variable.

When working hours were entered into the model as a continuous variable, we found an association between long hours and better phonemic fluency at baseline but not at follow-up. This inconsistency is also reflected in the lack of an association between the categorical working hours and phonemic fluency. More research is needed to determine whether employees with long working hours do better than other employees on tests of verbal fluency. Out of 15 analyses, we found 2 statistically significant interaction effects between working hours and sex, and sex-stratified analysis showed that long working hours were associated with poorer vocabulary performance among men but not among women. However, further research with larger samples is needed to examine potential sex differences in the association between working hours and cognition.

### Strengths and limitations

The strengths of this study include a large sample size and the possibility to explore prospectively the association between long working hours and a possible change in cogni-

tive function over a 5-year interval, which has not been feasible in earlier studies. Furthermore, we used 5 separate measures of cognitive function, allowing associations with specific aspects of cognition to be observed, and we were able to adjust for a large number of covariates as potential confounding or mediating factors between the exposure and outcome.

There are also important limitations in this study. First, the period of 5 years for cognitive decline might not be sufficient to detect a significant decline in cognitive function in general. Second, the Whitehall II cohort is based on civil servants and not representative of the entire working population, limiting the generalizability of our results. Third, we used self-reported working hours, with inherent problems of recall. Fourth, middle-aged occupational cohorts, such as ours, are subject to a healthy survivor effect as the study design involves participants who are employed and gradually excludes those who develop work disability. However, all cohort studies focusing on work-related exposures at midlife are open to health-related selection because participants need to be employed. Because poor health is linked with worse cognition, the healthy survivor effect is likely to lead to conservative estimates of the associations found. The baseline of the present study was approximately 15 years after inclusion into the Whitehall II study; men, employees in the higher occupational grades, and those free from coronary heart disease were slightly overrepresented. However, the associations among work hours, vocabulary, and reasoning were robust to adjustments for sex, occupational grade, and health. Furthermore, the similarity of these associations in the complete case and multiple imputation analyses suggests that loss to follow-up after the baseline is an unlikely source of bias in this study.

### Conclusions

Decline in cognitive function has already been shown to be present among the middle aged (9). As mild cognitive

impairment predicts dementia (10, 11) and mortality (56–58), the identification of risk factors for mild cognitive impairment in middle age is important. The results of this study show that long working hours may be one of the risk factors that have a negative effect on cognitive performance in middle age. Our findings can have clinical significance, as the 0.6- to 1.4-unit difference in aspects of cognitive functioning between employees working long hours and those working normal hours is similar in magnitude to that of smoking, a risk factor for dementia (59), which has been found to affect cognition in the Whitehall II study (60). However, further research is needed to identify the potential underlying factors for the relation between long working hours and cognitive function and to examine the generalizability of our findings.

### ACKNOWLEDGMENTS

Author affiliations: Centre of Expertise for Work Organizations, Finnish Institute of Occupational Health, Helsinki, Finland (Marianna Virtanen, Markus Jokela, Jussi Vahtera, Mika Kivimäki); Department of Epidemiology and Public Health, University College London, London, United Kingdom (Jane E. Ferrie, Archana Singh-Manoux, David Gimeno, Michael G. Marmot, Mika Kivimäki); INSERM, Saint-Maurice, Cédex, France (Archana Singh-Manoux); Centre de Gérontologie, Hôpital Ste Péline, Paris, France (Archana Singh-Manoux); Division of Environmental and Occupational Health Sciences, Health Science Center at Houston, The University of Texas School of Public Health, San Antonio, Texas, (David Gimeno); and the National Research and Development Centre for Welfare and Health, Helsinki, Finland (Marko Elovainio).

The Whitehall II study has been supported by grants from the British Medical Research Council; the British Heart Foundation; the British Health and Safety Executive; the British Department of Health; the US National Heart, Lung, and Blood Institute (grant HL36310); the US National Institute on Aging (grant AG13196); the US Agency for Health Care Policy and Research (grant HS06516); and the John D. and Catherine T. MacArthur Foundation Research Networks on Successful Midlife Development and Socioeconomic Status and Health. A. S.-M. is supported by a “European Young Investigator Award” from the European Science Foundation. M. G. M. is supported by a British Medical Research Council research professorship. J. E. F. is supported by the British Medical Research Council (grant G8802774). J. V. and M. K. are supported by the Academy of Finland (projects 117604, 124322, and 124271).

Conflict of interest: none declared.

### REFERENCES

- Vaguer C, Van Bastelaer A. Working overtime. In: *Statistics in Focus, Population and Social Conditions*. Dublin, Ireland: European Foundation for the Improvement of Living and Working Conditions; 2004.
- van der Hulst M. Long workhours and health. *Scand J Work Environ Health*. 2003;29(3):171–188.
- Caruso CC, Hitchcock EM, Dick RB, et al. *Overtime and Extended Work Shifts: Recent Findings on Illnesses, Injuries, and Health Behaviors*. Washington, DC: National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, Department of Health and Human Services; 2004.
- Johnson JV, Lipscomb J. Long working hours, occupational health and the changing nature of work organization. *Am J Ind Med*. 2006;49(11):921–929.
- Sokejima S, Kagamimori S. Working hours as a risk factor for acute myocardial infarction in Japan: case-control study. *BMJ*. 1998;317(7161):775–780.
- Liu Y, Tanaka H, the Fukuoka Heart Study Group. Overtime work, insufficient sleep, and risk of non-fatal acute myocardial infarction in Japanese men. *Occup Environ Med*. 2002;59(7):447–451.
- Sekine M, Chandola T, Martikainen P, et al. Work and family characteristics as determinants of socioeconomic and sex inequalities in sleep: the Japanese Civil Servants Study. *Sleep*. 2006;29(2):206–216.
- Shields M. Long working hours and health. *Health Rep*. 1999; 11(2):33–48.
- Kivipelto M, Ngandu T, Laatikainen T, et al. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol*. 2006;5(9):735–741.
- Chertkow H. Mild cognitive impairment. *Curr Opin Neurol*. 2002;15(4):401–407.
- Morris JC, Storandt M, Miller JP, et al. Mild cognitive impairment represents early-stage Alzheimer disease. *Arch Neurol*. 2001;58(3):397–405.
- Proctor SP, White RF, Robins TG, et al. Effect of overtime work on cognitive function in automotive workers. *Scand J Work Environ Health*. 1996;22(2):124–132.
- Lockley SW, Cronin JW, Evans EE, et al. Effect of reducing interns’ weekly work hours on sleep and attentional failures. *N Engl J Med*. 2004;351(18):1829–1837.
- Knauth P. Extended work periods. *Ind Health*. 2007;45(1): 125–136.
- Marmot M, Brunner E. Cohort profile: the Whitehall II study. *Int J Epidemiol*. 2005;34(2):251–256.
- Heim AW. *AH 4 Group Test of General Intelligence*. Windsor, United Kingdom: NFER-Nelson Publishing Company, Ltd; 1970.
- Raven JC. *Guide to Using the Mill Hill Vocabulary Scale With Progressive Matrices*. London, United Kingdom: HK Lewis; 1965.
- Borkowski JG, Benton AL, Spreen O. Word fluency and brain damage. *Neuropsychologia*. 1967;5(2):135–140.
- Launer LJ, Masaki K, Petrovitch H, et al. The association between midlife blood pressure levels and late-life cognitive function: the Honolulu-Asia Aging Study. *JAMA*. 1995;274(23):1846–1851.
- Breteler MM. Vascular involvement in cognitive decline and dementia. Epidemiologic evidence from the Rotterdam Study and the Rotterdam Scan Study. *Ann N Y Acad Sci*. 2000;903: 457–465.
- Singh-Manoux A, Sabia S, Lajnef M, et al. History of coronary heart disease and cognitive performance in midlife: the Whitehall II study. *Eur Heart J*. 2008;29(17):2100–2107.
- Swan G, Lessov-Schlaggar CN. The effects of tobacco smoke and nicotine on cognition and the brain. *Neuropsychol Rev*. 2007;17(3):259–273.
- Silvers JM, Tokunaga S, Berry RB, et al. Impairments in spatial learning and memory: ethanol, allopregnanolone, and

- the hippocampus. *Brain Res Brain Res Rev.* 2003;43(3):275–284.
24. Espeland MA, Gu L, Masaki KH, et al. Association between reported alcohol intake and cognition: results from the Women's Health Initiative Memory Study. *Am J Epidemiol.* 2005;161(3):228–238.
25. Britton A, Singh-Manoux A, Marmot M. Alcohol consumption and cognitive function in the Whitehall II study. *Am J Epidemiol.* 2004;160(3):240–247.
26. Colcombe S, Kramer AF. Fitness effects on the cognitive function of older adults: a meta-analytic study. *Psychol Sci.* 2003;14(2):125–130.
27. Bassuk SS, Berkman LF, Wypij D. Depressive symptomatology and incident cognitive decline in an elderly community sample. *Arch Gen Psychiatry.* 1998;55(12):1073–1081.
28. Yaffe K, Blackwell T, Gore R, et al. Depressive symptoms and cognitive decline in nondemented elderly women: a prospective study. *Arch Gen Psychiatry.* 1999;56(5):425–430.
29. Philibert I. Sleep loss and performance in residents and non-physicians: a meta-analytic examination. *Sleep.* 2005;28(11):1392–1402.
30. Singh-Manoux A, Richards M, Marmot M. Socioeconomic position across the lifecourse: how does it relate to cognitive function in mid-life? *Ann Epidemiol.* 2005;15(8):572–578.
31. Lee S, Buring JE, Cook NR, et al. The relation of education and income to cognitive function among professional women. *Neuroepidemiology.* 2006;26(2):93–101.
32. Koster A, Penninx BW, Bosma H, et al. Socioeconomic differences in cognitive decline and the role of biomedical factors. *Ann Epidemiol.* 2005;15(8):564–571.
33. Glymour MM, Weuve J, Fay ME, et al. Social ties and cognitive recovery after stroke: does social integration promote cognitive resilience? *Neuroepidemiology.* 2008;31(1):10–20.
34. Bassuk SS, Glass TA, Berkman LF. Social disengagement and incident cognitive decline in community-dwelling elderly persons. *Ann Intern Med.* 1999;131(3):165–173.
35. Lee S, Kawachi I, Grodstein F. Does caregiving stress affect cognitive function in older women? *J Nerv Ment Dis.* 2004;192(1):51–57.
36. Crowe M, Andel R, Pedersen NL, et al. Personality and risk of cognitive impairment 25 years later. *Psychol Aging.* 2006;21(3):573–580.
37. Lupien SJ, Maheu F, Tu M, et al. The effects of stress and stress hormones on human cognition: implications for the field of brain and cognition. *Brain Cogn.* 2007;65(3):209–237.
38. Potter GG, Helms MJ, Plassman BL. Associations of job demands and intelligence with cognitive performance among men in late life. *Neurology.* 2008;70(19 pt 2):1803–1808.
39. Ware JE Jr, Kosinski M, Bayliss MS, et al. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care.* 1995;33(4 suppl):AS264–AS279.
40. Kivimäki M, Head J, Ferrie JE, et al. Hypertension is not the link between job strain and coronary heart disease in the Whitehall II study. *Am J Hypertens.* 2007;20(11):1146–1153.
41. Goldberg DP. *The Detection of Psychiatric Illness by Questionnaire.* London, United Kingdom: Oxford University Press; 1972.
42. Stansfeld SA, Marmot MG. Social class and minor psychiatric disorder in British civil servants: a validated screening survey using the General Health Questionnaire. *Psychol Med.* 1992;22(3):739–749.
43. Jenkins R, Lewis G, Bebbington P, et al. The National Psychiatric Morbidity Surveys of Great Britain—initial findings from the household survey. *Psychol Med.* 1997;27(4):775–789.
44. Ferrie JE, Shipley MJ, Cappuccio FP, et al. A prospective study of change in sleep duration: associations with mortality in the Whitehall II cohort. *Sleep.* 2007;30(12):1659–1666.
45. Jenkins D, Stanton BA, Niemcryk S, et al. A scale for the estimation of sleep problems in clinical research. *J Clin Epidemiol.* 1988;41(4):313–321.
46. White I, Altmann DR, Nanchahal K. Mortality in England and Wales attributable to any drinking, drinking above sensible limits and drinking above lowest risk level. *Addiction.* 2004;99(6):749–756.
47. Kujala UM, Sarna S, Kaprio J, et al. Hospital care in later life among former world-class Finnish athletes. *JAMA.* 1996;276(3):216–220.
48. Stansfeld S, Marmot MG. Deriving a survey measure of social support: the reliability and validity of the close persons questionnaire. *Soc Sci Med.* 1992;35(8):1027–1035.
49. Karasek RA. Job demands, job decision latitude and mental strain: implications for job redesign. *Adm Sci Q.* 1979;24(2):285–308.
50. Glymour M, Weuve J, Berkman LF, et al. When is baseline adjustment useful in analyses of change? An example with education and cognitive function. *Am J Epidemiol.* 2005;162(3):267–278.
51. Royston P. Multiple imputation of missing values. *Stata J.* 2004;4(3):227–241.
52. Christensen H. What cognitive changes can be expected with normal ageing? *Aust N Z J Psychiatry.* 2001;35(6):768–775.
53. Blair C. How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behav Brain Sci.* 2006;29(2):109–160.
54. Lundberg U. Stress hormones in health and illness: the roles of work and gender. *Psychoneuroendocrinology.* 2005;30(10):1017–1021.
55. Garde AH, Faber A, Persson R, et al. Concentrations of cortisol, testosterone and glycosylated haemoglobin (HbA1c) among construction workers with 12-h workdays and extended workweeks. *Int Arch Occup Environ Health.* 2007;80(5):404–411.
56. Sabia S, Guéguen A, Marmot MG. Does cognition predict mortality in midlife? Results from the Whitehall II cohort study. *Neurobiol Aging.* (doi:10.1016/j.neurobiolaging.2008.05.007).
57. Pavlik VN, de Moraes SA, Szklo M, et al. Relation between cognitive function and mortality in middle-aged adults. The Atherosclerosis Risk in Communities Study. *Am J Epidemiol.* 2003;157(4):327–334.
58. Portin ML, Muuriaisniemi M, Joukamaa S, et al. Cognitive impairment and the 10-year survival probability of a normal 62-year-old population. *Scand J Psychol.* 2001;42(4):359–366.
59. Anstey KJ, von Sanden C, Salim A, et al. Smoking as a risk factor for dementia and cognitive decline: a meta-analysis of prospective studies. *Am J Epidemiol.* 2007;166(4):367–378.
60. Sabia S, Marmot M, Dufouil C, et al. Smoking history and cognitive function in middle-age from the Whitehall II study. *Arch Intern Med.* 2008;168(11):1165–1173.