Cognitive function following reported exposure to contaminated air on commercial aircraft: methodological considerations for future researchers

Sarah Mackenzie Ross,[1,*] Virginia Harrison,[2] Laura Madeley,[3] Kavus Davis,[4] Kelly Abraham-Smith,[5] Tessa Hughes[6] and Oliver Mason[1]

[1] *Research Department of Clinical, Educational and Health Psychology, University College London, Gower Street, London WC1E 6BT, UK*
[2] *Department of Psychology, Open University, Walton Hall, Milton Keynes MK7 6AA, UK*
[3] *Newham Diagnostic Memory Clinic, 103 First Avenue, Plaistow E13 8AP, UK*
[4] *North East London NHS Foundation Trust, Wood House, Thorpe Combe Hospital, 714 Forest Road, London E17 3HP, UK*
[5] *The School of Psychology, Doctorate in Clinical Psychology, University of Hertfordshire, Hatfield AL10 9AB, UK*
[6] *The School of Psychology, Doctorate in Clinical Psychology, Edwards Building, University of East London (Stratford Campus), Water Lane, London E15 4LZ, UK*

Cabin air on commercial aircraft is sometimes contaminated with jet engine oils containing organophosphates (OP). Aircrew have complained of chronic ill health and cognitive impairment following exposure to contaminated air, but a debate is ongoing about causation, diagnosis and treatment of long-term effects. The incidence of contaminated air events is difficult to quantify, as commercial aircraft do not have air quality monitoring systems on board. According to statistical records, certain types of aircraft suffer more fume events than others (e.g. the BAe 146 and Boeing 757) and it has been suggested that airframe may serve as a proxy measure of exposure. The current study sought to investigate this claim, and to determine whether an association exists between exposure to contaminated air and neuropsychological impairment. Twenty-nine pilots were recruited and split into two exposure groups according to aircraft type flown, but few differences were noted between groups in terms of exposure history or cognitive function. Pilots' profile of cognitive performance deviates from that seen in the normal population, but mirrors that seen in other OP-exposed cohorts. In particular they show decrements in performance on tests of attention, psychomotor speed and visual sequencing. Given the safety implications of these findings, further research is warranted.

Keywords: aviation, neurobehaviour, neuropsychology, neurotoxicology, organophosphates, pilots

1. INTRODUCTION

Cabin air on commercial aircraft is sometimes contaminated with hydraulic fluids, synthetic jet engine oil and combusted or pyrolysed materials. The process by which this happens is as follows: In civil jet aircraft, outside air is bled off the gas turbine engine or auxiliary power unit to pressurize the cabin and provide breathing air. It is unfiltered and, occasionally, faulty seals allow engine oil fumes to enter the cabin. Contaminated air may contain a large number of chemicals that can cause irritation, skin sensitization, and respiratory and neurological problems [1–3]. Any aircraft that relies on bleed air to ventilate the cabin can in theory develop faulty engine seals and suffer occasional engine oil leaks, but for certain types of aircraft more contaminated air events are statistically recorded than for others, in particular the BAe 146 and Boeing 757 [4].

* Corresponding author. E-mail: s.mackenzie-ross@ucl.ac.uk Tel.: +44 207 679 5699. Fax: +44 207 916 1989.

The incidence of contaminated air events on commercial aircraft is difficult to quantify as commercial aircraft do not have air quality monitoring systems on board. Underreporting of contaminated air events is common amongst aircrew due to lack of awareness, commercial pressure and fears over job security if crew complain about working conditions; many crews see contaminated air as a normal, everyday occurrence [4, 5].

Aircrew and some passengers around the world have been reporting ill health following contaminated air events for many years [6–12] but it is only recently that this issue has received attention in the UK [12–14]. The immediate effects of exposure to contaminated air have been well documented and include eye irritation, respiratory problems, headache, skin problems, nausea, vertigo, loss of balance, dizziness, fatigue and cognitive impairment (disorientation, confusion and memory problems). These symptoms show a close temporal relationship with exposure and usually recede after cessation of contact [5].

Journal of Biological Physics and Chemistry **11** (2011) 180–191
Received 11 October 2011; *accepted* 30 December 2011

doi: 10.4024/30MA11A.jbpc.11.04

A number of individuals report persistent, chronic ill health lasting months or years after exposure including cognitive impairment (memory, word finding, multitasking difficulties), lack of coordination, nausea/vomiting, diarrhoea, respiratory problems, chest pains, severe headaches, light-headedness, dizziness, weakness and fatigue, paraesthesias, tremors, increased heart rate, palpitations, irritation of ear, nose and throat, muscle weakness/pain, joint pain, salivation, skin itching, rashes, blisters, hair loss, signs of immunosuppression and chemical sensitivity [5–12].

A debate is ongoing in the UK about causation, diagnosis and treatment of long term effects [13, 15, 16]. Aircrew are concerned that these persistent symptoms may be associated with exposure to an organophosphate (tricresyl phosphate, TCP), which is used as an antiwear additive in jet engine lubricants and hydraulic fluids [17–19]. This chemical is a known neurotoxin thought to be responsible for the paralysis seen in thousands of American individuals in the 1930s who consumed a drink called "Ginger Jake", which had been contaminated [20, 21]. Until very recently, no biomarker of past cumulative exposure to TCP has been available, which might allow links to be drawn between prior exposure and chronic health effects.[1] However, the potential for exposure has been documented by air quality studies undertaken during routine flights. Cabin air contamination by TCP in commercial aircraft has been documented by two previous studies [22, 23]. However, the concentration of TOCP (tri-*ortho*-cresyl phosphate) found in cabin air by Crump et al. [22] was described as being within safe limits, despite the fact that exposure standards for TCP in aircraft do not currently exist. Furthermore, it is important to note that Crump et al.'s air quality study was undertaken during routine flights when no fume events triggering the airline's formal reporting procedure occurred. To date, no monitoring of cabin air has ever been successfully undertaken during a contaminated air event. Therefore, the nature of the contaminants within the cabin air and the levels of exposure to passengers and crews during reportable contaminated air events are unknown.

In a previous paper we reported the results of neuropsychological assessment of 27 UK airline pilots who complained of ill health and cognitive impairment following exposure to contaminated air [14]. Pilots had either been referred by other medical specialists or referred themselves for psychological assessment as they were concerned about their cognitive function and capacity to fly. All except one were current or former pilots on the Boeing 757 or BAe 146 aircraft types. They were assessed by multiple examiners instructed to seek explanations other than exposure to toxic fumes for any symptoms identified during the assessment. Nine pilots were subsequently excluded from group analysis of cognitive function as they had a medical or psychiatric history that might otherwise explain their symptoms of ill health. Psychometric assessment of the remaining 18 pilots revealed language, perceptual and general intellectual ability were preserved but performance on tests of psychomotor speed attention and executive functioning was below expected levels. Indeed, the pilots reported alarming cognitive failures at work, such as being unable to retain numerical information from air traffic control. Nine were still flying, 4 were on sick leave and 5 had retired from the profession. The cognitive deficits identified in this study were not attributable to mood disorder or malingering, but firm conclusions regarding the causal link with contaminated air could not be drawn as only crude measures of exposure were available (i.e., based on self-reporting) and the sample of pilots was small and self-selected, making it impossible to determine how representative they are of the broader community of commercial airline pilots.

The first author recommended that the UK Government commission further research to establish the prevalence of ill health amongst aircrew and the relationship, if any, to working practices and exposure to contaminated air. The first author highlighted the importance of recruiting a control group of pilots who have not been exposed to contaminated air to establish whether the profile of cognitive deficits noted in her case series is common amongst all pilots. If so, it would suggest that lifestyle factors or some other aspects of flying are responsible for ill health rather than exposure to contaminated air.

The first author presented her findings and recommendations to a UK Government scientific advisory panel called the Committee on Toxicity of Chemicals in Food Consumer Products and the Environment (COT). This committee had been asked by the UK Government Department of Transport (DfT) to review the available scientific evidence concerning the possible effects on aircrew health of exposure to oil/hydraulic fluid smoke/fumes. The COT met several times during 2006 and 2007 to discuss this issue and concluded in 2007 that the evidence to date does not allow firm conclusions to be drawn regarding a link between exposure to contaminated

[1] Researchers from the Universities of Washington and Nebraska recently developed a biomarker of exposure to TCP which can determine whether an individual has been exposed to TCP during the last month. However, it cannot reliably detect exposures that occurred more than one month prior to testing [17–19].

air and the development of chronic ill health. They recommended that the DfT commission further research, in particular:

(1) Onboard air quality monitoring studies (COT [13], statement paragraphs 65–72 and 87–91);

(2) A cross-sectional epidemiological study of neuropsychological outcome, incorporating multidisciplinary and international collaboration. The COT acknowledged accurate exposure assessment would be difficult given the absence of routine onboard air quality monitoring on commercial aircraft, rendering it impossible to determine what chemicals enter the cabin or in what quantities. The COT proposed an alternative approach: to explore performance on neuropsychological tests in relation to proxy measures of exposure such as airframe type flown. They suggested that future researchers investigate whether performance in neuropsychological tests differs between pilots who fly different airframes (COT statement [13], paragraphs 82–83 and 94–95).

The first author contacted the DfT to apply for research funding but was advised that an epidemiological study would not be commissioned until the results of the air quality monitoring study were available. Researchers from Cranfield University were commissioned in 2008 to undertake the air quality work, but the findings were only published recently [22]. In the interim, this paper reports a small scale neuropsychological study (unfunded) of a random sample of working pilots aimed at establishing whether pilots who fly aircraft types associated with contaminated air events show evidence of cognitive impairment and to determine whether the profile of deficits is similar to that seen in the initial self-selected sample of 18 pilots previously described by Mackenzie Ross et al. (2008) [14]. Attempts were made to recruit a sample of nominally unexposed pilots in order to determine whether any cognitive impairment identified in exposed pilots is associated with exposure to contaminated air or some other aspect of flying such as jet lag, shift work or exposure to ozone.

2. METHOD

2.1 Ethics

Ethical approval for this study was granted by University College London research ethics committee. Written informed consent was obtained from all study participants.

2.2 Specific hypotheses

(1) If airframe is an appropriate proxy measure of exposure then pilots who fly airframes which record more contaminated air events than others (i.e., the BAe 146 and Boeing 757) will report more fume events than pilots who fly other types of aircraft.

(2) If an association exists between exposure to contaminated air and the development of neuropsychological impairment, then pilots who fly airframes associated with contaminated air events will show evidence of cognitive impairment. In contrast, pilots who fly other aircraft types will not show evidence of cognitive impairment or else will exhibit a different profile of cognitive strengths and weaknesses compared to exposed pilots. If unexposed and exposed pilots show a similar pattern of cognitive impairment, this suggests some other aspect of flying or common lifestyle factor is responsible for ill health.

(3) If an association exists between exposure to contaminated air and the development of neuropsychological impairment, then pilots who fly aircraft types associated with contaminated air events will show the same pattern of cognitive impairment as seen in the self-selected sample of 18 pilots previously described by Mackenzie Ross et al. (2008). In particular, they will show reduced performance on tests of psychomotor speed attention and executive functioning, although levels of impairment are likely to be lower than that seen in the initial group of 18 pilots as half of the latter were too ill to continue flying.

2.3 Study participants

The identification and recruitment of suitable study participants involved collaboration with other organizations. We contacted the Independent Pilots Association (IPA), which represents 1500 pilots, and they agreed to contact their members on our behalf to ask them if they would assist us with our study.

2.4 Study design and exclusion criteria

The study was undertaken in two phases. The first phase involved contacting all 1500 pilots on the IPA database to ask for their contact details, work history and willingness to take part in our study. The second phase involved asking a random subsample of the respondents to undergo neuropsychological assessment so that the performance of pilots who fly aircraft types associated with contaminated air events could be compared with pilots who fly other aircraft types that are not associated with contaminated air events (control group). Pilots were excluded if they had a history of substance abuse (including alcohol) or a history of psychiatric, neurological or serious medical problems that might affect performance on psychometric tests. A total of 29 pilots underwent neuropsychological assessment.

2.5 Measures

Work history and physical symptoms. A questionnaire was devised concerning demographic, work and exposure history.

Mood State. The Hospital Anxiety and Depression Scale was used to assess symptoms of distress [24].

Neuropsychological assessment. In order to test the possible effects of organophosphate exposure on cognitive functioning, an extensive test battery was carried out utilizing tests that are known to be reliable, sensitive to impairments and routinely used in clinical practice in the UK. This battery, detailed below, included psychometric tests of response speed, executive function, working memory, verbal ability, visual and auditory memory and fine motor control, all of which have adequate published reliability, validity and normative data.

The Wechsler Adult Intelligence Scale (WAIS)-III was used to measure current intellectual functioning. It comprises fourteen subtests, which measure a variety of verbal and nonverbal functions, working memory and processing speed. Eleven subtests were administered [25].

The Adult Memory and Information Processing Battery was used to assess verbal and nonverbal memory, and mental and motor processing speed [26]. It comprises six subtests.

Trail Making A & B were used to assess motor speed and mental flexibility and the Stroop test was included as a measure of mental flexibility [27, 28].

A verbal and semantic fluency test [29] was used to assess executive function.

The California Computerized Assessment Package was used to assess simple and choice reaction time [30].

The Medical Symptom Validity Test was used as a measure of cognitive effort/symptom validity [31]. It is a brief computerized verbal memory screening test that measures a person's effort on testing and was included in the battery to ensure psychometric test results were valid. It is insensitive to all but the most extreme forms of cognitive impairment whilst being very sensitive to poor effort and exaggeration of cognitive difficulties.

3. RESULTS

3.1 Recruitment rates

Only 70/1500 questionnaires were returned. This is a very low proportion of the total sample (4.7%) and it is unclear how representative these pilots are of the broader community of airline pilots. A random subsample of 29 pilots was asked to undergo neuropsychological assessment. Study participants were split into two groups using proxy measures of exposure (i.e., aircraft type flown): an "exposed" group consisting of pilots who flew the BAe 146 or the Boeing 757 ($n = 15$) and an "unexposed" group consisting of pilots who flew other aircraft types such as the Boeing 737 ($n = 14$).

3.2 Demographic information

Measures of age, years in education, years spent flying, current and premorbid intelligence were obtained for the two groups of pilots (see Table 1). Commonly used tests of premorbid ability estimate IQ from reading skill (e.g., the National Adult Reading Test and the Wechsler Test of Adult Reading). However, reading tests may underestimate premorbid IQ in individuals with a history of exposure to OPs as previous research suggests organophosphate-exposed individuals may have impaired reading ability [32]. Therefore, in the current study premorbid intelligence was estimated using a measure that is unlikely to have been affected by cognitive damage (matrix reasoning). Independent *t*-tests revealed that the groups were successfully matched for age, years flying, premorbid IQ (all $t < 1$) and current IQ ($t(27) = -1.37$, *ns*[2]); and Mann Whitney *U* tests confirmed matching was successful for education ($U = 99.00$, *ns*).

Table 1. Demographic information for the two pilot groups.

	Exposed group		Unexposed group	
	Mean	(SD)	Mean	(SD)
Age	50.4	(9.4)	52.1	(9.2)
Years of education	14.3	(2.0)	14.6	(2.2)
Years flying	20.0	(13.2)	22.6	(14.4)
Matrix reasoning	15.1	(1.8)	14.9	(2.5)
WAIS-III full scale IQ	138.7	(8.9)	134.1	(9.2)

3.3 Exposure Information

Accurate estimation of exposure is critical for the validity of studies investigating the health effects of occupational exposures to toxic substances. As mentioned previously, on-board monitoring of air quality is not routinely undertaken on commercial aircraft, therefore the nature of any contaminants within the cabin air and the levels of exposure to passengers and crews during a fume event is unknown.

Furthermore, there are no routine laboratory tests available to determine whether an individual has been exposed to TCP during flight and so the present study investigated the reliability of using a simple proxy measure of exposure—airframe flown—in order to classify participants into "exposed" and "unexposed" groups. This approach assumes that pilots who fly specified aircraft types associated with contaminated air events (e.g., the BAe 146 and Boeing 757) will have a history of exposure to toxic fumes, whilst pilots who fly other aircraft types will not have been exposed to toxic fumes during their career. In order to establish if this is the case, pilots were

[2] not significant.

interviewed about their work and exposure history and the information gleaned appears in Figure 1 and Table 2.

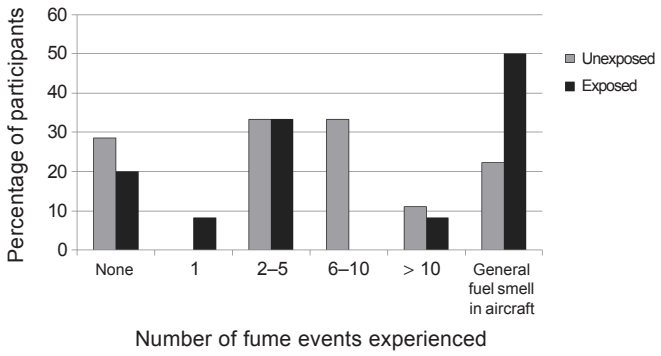

Figure 1. Frequency of fume events experienced by group.

Table 2. Proportions of pilots from each group who had experienced a fume event.

	Unexposed group %	Exposed group %
Percentage of participants who experienced at least one fume event	64	80
Of those who experienced a fume event, how many reported it?	67	83
Of those, who reported it, how many events were confirmed by an aircraft engineer?	50	100
Of those who experienced a fume event, how many felt it lead to ill health?	22	33
Of those who felt unwell, how many sought medical help?	0	50

As illustrated above, the majority of pilots in both groups reported having experienced at least one fume event at some point in their career, with most participants reporting multiple exposures. Only 7 pilots stated they had never experienced a fume event and they were equally distributed across the two groups. Fisher's Exact Test analysis revealed no significant association between pilot group and whether the pilots reported fume event experience. Based on this evidence, aircraft type flown may not be a reliable indicator of exposure history and is likely to result in exposure misclassification.

3.4 Cognitive performance

The performance of exposed and control cohorts in neuropsychological tests can be seen in Table 3.

3.4.1 Exposed versus supposedly "unexposed" cohort comparisons

Differences in performance between exposed and supposedly unexposed cohorts on psychometric testing were analysed using parametric and nonparametric statistical tests. Where possible independent *t*-tests were used; however, where the assumption of normality was violated, Mann Whitney *U* tests were conducted. All analyses were Bonferroni corrected *within* their family of tests. For example, pilots' level of intellectual functioning was assessed using WAIS-III. As this test comprises 11 subtests, a series of Bonferroni-corrected independent *t*-tests (using a new α of 0.0045) were conducted with exposure group as the independent variable. No significant differences between the two groups' performance on any of the WAIS-III subtests were found (largest $t = 3.06$). Similarly, no significant differences were identified between the two groups on any of the other neuropsychological tests or mood state measures.

These findings suggest that the two groups have a similar profile of neuropsychological performance. However, given the fact that there were no significant group differences in fume event exposure, this may not be surprising. Instead, it may be that using airframe as a proxy measure of exposure is inappropriate, and as such our control group is, in fact, another exposed group of pilots.

Therefore, the two pilots groups were collapsed into one group comprised only of pilots who report a history of exposure to contaminated air events. This left a sample size of $n = 22$ after excluding the seven previously described pilots who stated that they had never experienced a fume event. The exposed pilots' performance on neuropsychological tests was compared to published test norms derived from a cross section of healthy adults in the general population.

3.4.2 Pilots versus normative comparison standards: reanalysis of the data using an alternative comparison group

The WAIS-III test of intelligence has been developed over many years and is the result of extensive empirical studies in the US and UK involving a standardization sample of over 2000 adults aged 16–90 years. The WAIS-III comprises 14 subtests, which measure a range of abilities thought to comprise intelligence, such as working memory, processing speed and verbal and visual reasoning. Contemporary normative information and interpretive tables are provided to allow an individual's performance on the WAIS-III to be compared to national norms.

Often an individual's performance varies across the different subtests that make up WAIS-III and this is particularly true of high ability subjects. WAIS-III not only provides an overall full scale IQ score and separate index scores for broad cognitive domains such as verbal and visual ability, working memory and processing speed; it

Table 3. Mean subtest scores of pilots from both groups.

Test groups	Subtests	Unexposed group		Exposed group		Test statistics	
		Mean	*SD*	Mean	*SD*		
WAIS-III subtests	Information	14.71	*1.82*	15.07	*2.05*	*t*	–0.49
	Digit Span	14.36	*2.47*	14.13	*2.70*	*t*	0.23
	Vocabulary	14.86	*2.88*	15.80	*1.32*	*t*	–1.12
	Arithmetic	14.71	*2.40*	15.07	*1.53*	*t*	–0.48
	Comprehension	15.86	*2.38*	17.13	*1.55*	*t*	–1.72
	Similarities	12.93	*2.95*	15.33	*1.95*	*t*	–2.61
	Picture Completion	13.71	*1.94*	15.87	*1.85*	*t*	–3.06
	Picture Arrangement	13.29	*2.95*	13.20	*2.14*	*t*	0.09
	Block Design	15.00	*2.63*	15.27	*2.15*	*t*	–0.30
	Matrix Reasoning	14.86	*2.48*	15.13	*1.81*	*t*	–0.35
	Digit Symbol	13.21	*3.49*	11.80	*2.21*	*t*	1.31
AMIPB visual memory subtests	Complex Figure Immediate	95.00	*4.31*	93.93	*5.22*	*t*	0.60
	Complex Figure Delayed	93.79	*7.00*	93.52	*5.50*	*t*	0.12
	Complex Figure Retained	98.50	*4.88*	98.99	*2.37*	*U*	90.0
	Design Learning Trials 1–5	40.07	*3.83*	39.80	*4.95*	*t*	0.16
	Design Learning Recall	8.64	*0.63*	8.33	*1.50*	*t*	0.72
AMIPB verbal memory subtests	List Learning Trials 1–5	54.50	*8.79*	52.67	*9.02*	*t*	0.55
	List Learning Free Recall	11.86	*2.63*	11.20	*2.65*	*t*	0.67
	List Learning-Recognition	14.14	*1.29*	14.27	*0.80*	*U*	108.5
	Story: Immediate Recall	42.50	*6.30*	46.87	*4.81*	*t*	–2.11
	Story: Delayed Recall	40.71	*6.68*	43.47	*5.84*	*t*	–1.18
	Story: Retained	95.71	*7.55*	92.76	*4.79*	*t*	1.27
Motor speed and mental flexibility tests	Trails A Score	23.00	*7.18*	26.53	*10.41*	*t*	–1.06
	Trails B Score	52.43	*25.73*	54.87	*19.38*	*U*	128.5
	Stroop Score	107.21	*9.96*	111.93	*0.26*	*U*	144.5
Executive function tests	Verbal Fluency	51.71	*6.41*	49.73	*11.23*	*t*	0.58
	Semantic Fluency	24.00	*3.62*	26.33	*3.89*	*t*	–1.67
Reaction time tests	CALCAP Simple	0.14	*0.53*	0.00	*0.00*	*t*	1.04
	CALCAP Complex	0.31	*0.63*	0.15	*0.55*	*t*	0.66
Mood state	HADS Anxiety	3.93	*2.53*	4.40	*3.40*	*t*	–0.42
	HADS Depression	2.64	*2.24*	3.47	*2.75*	*t*	–0.88

also allows evaluation of the variability in performance across all 14 subtests and calculates whether any observed differences in subtest performance are statistically significant. The procedure for evaluating an individual's cognitive strengths and weaknesses involves comparing a single subtest score to an average score on a group of subtests. This is referred to as discrepancy analysis. Interpretive tables enable the examiner to determine whether any variability in performance is statistically significant and clinically meaningful.

Figure 2 depicts the pattern of performance of the exposed pilots ($n = 22$) on the different WAIS-III subtests in

comparison to published test norms. The first thing to note is that the pilots obtained higher scores on all WAIS-III subtests compared to the UK standardization sample (t range = 4.26–13.82 with smallest $P < 0.0005$). Their overall IQ is higher than the UK mean, which is perhaps not surprising given the nature of their occupation. However, of particular interest is the fact that the exposed pilot group exhibits a profile of performance which is quite different from that seen in the normal population.

WAIS-III discrepancy analysis revealed that 32% of the pilots performed significantly lower than expected on a test of psychomotor speed (digit symbol); and 18% of pilots performed below expected levels on tests of attention (digit span), verbal reasoning (similarities) and visual sequencing (picture arrangement). This is consistent with pilots' subjective complaints of attentional problems and mental slowing. Overall these findings suggest that, compared to the normal UK population, the pilots exhibit significant patchy underfunctioning on certain WAIS-III subtests. This profile of subtest performance deviates from that seen in the normal population and may be indicative of a subtle, yet specific, pattern of cognitive deficits.

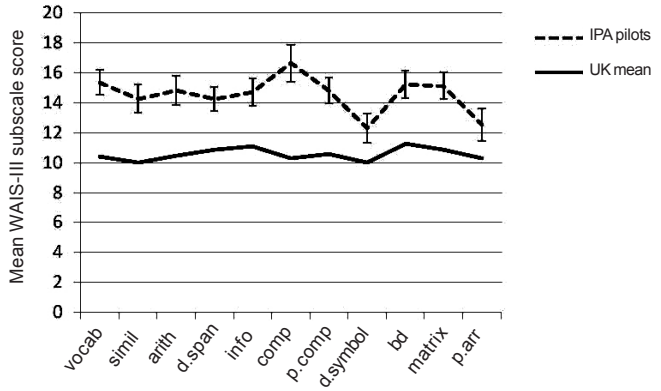

Figure 2. Performance profiles on WAIS-III subtests (error bars represent ± 2 S.E.), UK data taken from Wycherley et al. [33]. Key: vocab, vocabulary subtest; simil, similarities; arith, arithmetic; d.span, digit span; info, information; comp, comprehension; p.comp, picture completion; d.symbol, digit symbol; bd, block design; matrix, matrix reasoning; p.arr, picture arrangement.

3.4.3 Profile analysis

Neuropsychologists often diagnose cerebral dysfunction when patients show marked discrepancies amongst test scores [34]. Although a degree of intraindividual variation in test performance is normal, what is of interest is whether the profile of performance is recognizable and associated with specified medical disorders [35]. Of particular interest is whether the profile of performance seen in pilots is equivalent to that seen in other populations exposed to organophosphates.

In order to determine whether the pattern of cognitive performance seen in the pilots is related to organophosphate (OP) exposure, their performance was compared to that of two other cohorts who report ill health following exposure to OP compounds. The first group was the self-selected sample of 18 pilots previously reported by Mackenzie Ross [14]; while the other comprised farmers who had been exposed to organophosphate sheep dips, also previously described by Mackenzie Ross et al. [36]. Previous research with these groups has found patchy underfunctioning in certain cognitive domains, which may be attributable to organophosphates. Figure 3 directly compares the current pilot group to the previously seen exposed cohorts. Demographic information is given in Table 4 for comparison purposes.

Table 4. Demographic information for organophosphate (OP) exposed groups.

	Current IPA Pilots		SMR (2008) Pilots		SMR (2010) Farmers	
	Mean	(SD)	Mean	(SD)	Mean	(SD)
Age	50.5	(7.7)	48.4	(8.8)	54.7	(9.4)
Years of education	14.6	(2.2)	13.2	(2.4)	11.6	(2.1)
Matrix reasoning	15.1	(2.1)	14.4	(2.0)	12.4	(2.5)
WAIS IQ	136.3	(8.1)	119.3	(10.5)	104.8	(11.8)

Pilots perform better than the other two groups on all subtests of the WAIS-III, which would be expected as they have higher overall IQs (see Figure 3). However, their profile of performance mirrors that of the other OP-exposed cohorts. All show pronounced dips in performance on tests of attention (digit span), psychomotor speed (digit symbol) and visual sequencing (picture arrangement).

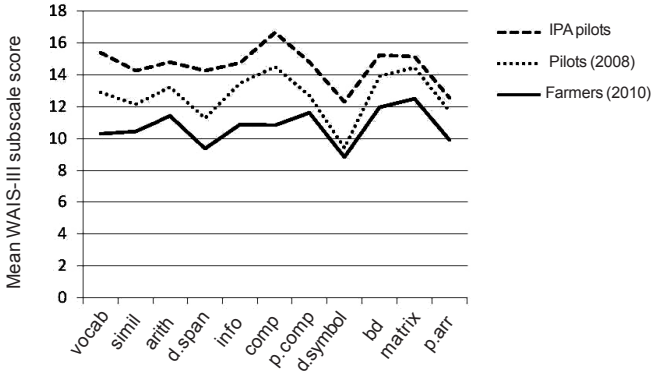

Figure 3. Performance profile comparisons to OP-exposed cohorts on WAIS-III subtests. Key: vocab, vocabulary subtest; simil, similarities; arith, arithmetic; d.span, digit span; info, information; comp, comprehension; p.comp, picture completion; d.symbol, digit symbol; bd, block design; matrix, matrix reasoning; p.arr, picture arrangement.

3.4.4 Relationship between self-reported exposure and cognitive function

Further analyses were undertaken to investigate the relationship between pilots' self-reported exposure to fume events (i.e., experiencing 0, 1, 2–5, 6–10, or > 10 events[3] as seen in Figure 1) and performance in tests of attention, psychomotor speed and visual sequencing. To increase the power of this analysis, data from the current study was combined with that collected previously by Mackenzie Ross et al. [14], giving a final sample size of 47 pilots.[4] A series of 1-tailed Spearman's correlations were undertaken and the results revealed significant negative correlations between exposure level and digit span ($R_s = –0.36$, $P < 0.05$) and picture arrangement ($R_s = –0.50$, $P < 0.005$). A moderate correlation was also noted with digit symbol ($R_s = –0.24$, $P < 0.10$). Only the correlation between exposure and picture arrangement remained significant following Bonferroni correction for type 1 error (new $\alpha = 0.016$).

4. DISCUSSION

The primary aim of the current study was to determine whether exposure to contaminated air on commercial aircraft is associated with neuropsychological impairment. The study design sought to address methodological weakness of earlier work by: (1) recruiting a random selection of pilots who had been exposed to contaminated air (rather than self-selected samples as previously reported in the literature [7, 14]); and (2) recruiting a cohort of pilots who had *not* been exposed to contaminated air to act as a control group. This was to enable researchers to determine the prevalence of neuropsychological problems amongst a representative sample of pilots and to establish whether any cognitive deficits identified in the study relate to exposure to contaminated air or some other aspect of flying such as shift work, jet lag or exposure to ozone. Unfortunately a number of methodological problems arose during the course of the study that complicate interpretation of the data. These include recruitment difficulties, sample size/bias and difficulty establishing a reliable estimate of exposure.

4.1 Exposure assessment

Accurate estimation of exposure is critical for the validity of studies investigating causal links between neuropsychological impairment and exposure to neurotoxic substances. However, objective evidence of exposure, such as environmental monitoring data or biomarkers of exposure, is seldom available in occupational studies [37]. This is particularly true in commercial aviation, where on-board monitoring of cabin air quality is not routinely undertaken. Previous research investigating the relationship between exposure to contaminated air on commercial aircraft and the development of chronic ill health has relied upon rough estimates of exposure based on an individual's testimony regarding their exposure history. Since individuals differ in their capacity to detect noxious substances and given the limits of human memory, exposure information collected in this way may be unreliable.

The present study used a proxy measure of exposure, namely aircraft type flown. This approach had been suggested by the Committee on Toxicity (COT), a UK Government scientific advisory panel, which reviewed the available scientific evidence concerning the possible effects on aircrew health of exposure to oil/hydraulic fluid-derived smoke/fumes [13]. The COT noted that certain aircraft types (i.e., BAe 146, Boeing 757) record statistically more contaminated air events than other aircraft types and suggested future researchers investigate whether performance on neuropsychological tests differs between pilots who fly different airframes.

The present study recruited a random sample of 29 pilots and split them into two groups according to aircraft type flown: An "exposed" group consisting of pilots who flew the BAe 146 or the Boeing 757 and an "unexposed" group consisting of pilots who flew other aircraft types such as the Boeing 737. Surprisingly, few differences were noted between exposed and supposedly unexposed cohorts in terms of exposure history or performance on neuropsychological tests. However, the majority of pilots in both groups reported experiencing at least one fume event at some point in their career, with most participants reporting multiple exposures, indicating that aircraft type flown is unlikely to serve as an accurate proxy measure of exposure. Furthermore, both groups of pilots obtained similar scores on neuropsychological tests, which is perhaps unsurprising as there were actually no significant differences between the groups in terms of exposure history.

Since the proxy measure of exposure used in this study resulted in exposure misclassification, the data were analysed in a different way. Both pilot groups were treated as having been potentially exposed to contaminated air and their performance in neuropsychological tests was compared with that seen in a UK standardization sample of over 2000 healthy adults. When compared to members

[3] Those who reported general fuel smell in the aircraft were excluded from this analysis, as it unclear at this stage whether this constitutes a greater or lesser amount of exposure compared to those who have reported exposure to specific fume events.
[4] The two studies utilized the same testing procedures.

of the general population, the pilots exhibited pronounced decrements in performance on tests of attention (digit span), psychomotor speed (digit symbol) and visuo-spatial sequencing (picture arrangement). This was consistent with their subjective complaints of attentional problems and mental slowing in everyday life. Overall these findings suggest the performance profile in both groups of pilots deviates from that seen in the normal population and may reflect a subtle, specific pattern of cognitive deficit.

To investigate this possibility further, the pattern of cognitive performance seen in both pilot groups was compared to that seen in two other cohorts who report ill health following exposure to OP compounds; the self-selected sample of pilots who had been exposed to contaminated air, previously described by Mackenzie Ross in 2008 [14] and a random sample of farmers exposed to pesticides containing OPs, described by Mackenzie Ross et al. in 2010 [36]. Both occupational groups show pronounced performance dips, suggesting that pilots who have been exposed to contaminated air show a similar profile of cognitive deficits as that seen in other occupational groups who have been exposed to OPs.

4.2 Sample bias and size

Unfortunately further methodological weaknesses limit the conclusions that can be drawn from this study. Population characteristics are usually inferred from measures taken from samples. If a sample is not truly representative of the population from which it is drawn then it is impossible to make accurate predictions about the population as a whole. It remains unclear how representative the sample of pilots recruited in the current study are of the broader community of pilots in the UK.

Identification and recruitment of UK airline pilots requires collaboration with other organizations. Over ten thousand pilots are listed on trade union databases, but the UK's largest union, which represents over 10,000 pilots, was unwilling to assist us with recruitment. A smaller union agreed to contact members on our behalf but has only 1500 members on their list. The response rate to our appeal for study participants was incredibly low (< 5%) and this, coupled with the fact the current study was unfunded and hence limited in scope, meant only a small number was examined. It is therefore difficult to determine how representative our final sample was of the airline industry as a whole. We were unsuccessful in identifying a sample of pilots who had not been exposed to contaminated air, making it impossible to determine the prevalence of ill health in exposed and unexposed populations and to unequivocally establish links, if any, with exposure history versus other aspects of flying.

4.3 Study strengths

Although the current study suffers from the weaknesses described in §§4.1 and 4.2, it also has several strengths. It involved detailed neuropsychological assessment, which is considered by many researchers to be the most sensitive means of examining the effects of neurotoxic exposure as it reveals more regarding subclinical effects than internal dose indicators such as levels of toxins in blood or urine [35, 37, 38]. Indeed, many toxins are metabolized in the human body and excreted quickly and may not leave biological markers to prove exposure or allow the level of exposure to be determined. Hence, neuropsychological testing is a useful diagnostic tool in the assessment of exposed persons. This study allowed the nature and extent of neurobehavioural problems to be explored in considerable depth, using clinically sensitive measures rather than administering brief screening tests that likely lack sensitivity and/or specificity. The psychometric test battery was designed to cover a range of cognitive functions and included tests that are routinely used in clinical practice for diagnostic purposes and for which reliable and valid normative data is available. Pilots were found to have deficits in particular areas whilst other abilities appeared intact, an important finding that needs further exploration.

This study took account of a number of potentially confounding variables and excluded individuals with a past medical and psychiatric history that could otherwise account for ill health. Current mood state was assessed because depression and anxiety can affect performance on psychometric tests and measures of effort were included to determine whether low motivation or deliberate exaggeration of symptoms influenced test results; however, no difference was found between the two groups on these measures. Hence, it is unlikely that any of these variables could account for the cognitive deficits identified in the pilots.

4.4 Implications

This study found reduced performance on specific cognitive tests in a random sample of UK airline pilots, all of whom were flying at the time of assessment. Study limitations make it impossible to establish or rule out a link between the abnormalities detected and exposure to contaminated air. In order to determine whether such a link exists a large scale epidemiological study should be undertaken to establish the prevalence of ill health (physical and psychological) amongst aircrew and the relationship, if any, with working practices and exposure to contaminated air. This study highlights methodological considerations for future researchers: identification and

recruitment of suitable study participants; selection of an appropriate control group (which ideally would comprise unexposed pilots, if such a group exists); difficulties inherent in exposure assessment; and selection of appropriate, sensitive outcome measures.

In 2007 the COT recommended that the UK Department for Transport (DfT) commission a cross-sectional epidemiological study of neuropsychological outcome, incorporating multidisciplinary and international collaboration. The first author was advised by the DfT that an epidemiological study would not be commissioned until the results of an air quality monitoring study, commissioned in 2008, were available. This study was undertaken by researchers from Cranfield University and the results were made public earlier this year [22]. The principle objective of the study was to measure levels of volatile organic compounds, semivolatile organic compounds and carbon monoxide in aircraft cabin air during various phases of flight and during a fume event. A total of 100 flights in 5 different aircraft types were monitored (Boeing 757, BAe 146, Airbus 320/1 and 319). Cabin crew were asked to complete a questionnaire at the end of each flight to determine whether they noticed any unusual smells. Fumes were reported on 38 flights but not at levels triggering formal reporting procedures.

Various chemicals were detected in the cabin air during routine flight, including the organophosphate tricresyl phosphate (TCP), carbon monoxide and toluene, but levels were reported to be within safety limits. With regard to TCP, aircraft safety standards do not exist, so the Cranfield researchers referred to other guidelines (e.g., indoor environmental standards for tri-*ortho*-cresyl phosphate) and concluded from them that levels within the aircraft cabin did not exceed available health and safety standards. It therefore seems unlikely that the DfT will commission epidemiological research, given Cranfield's findings.

However, the Cranfield study has been criticized on a number of grounds [39, 40]. Airborne concentrations of pollutants entering the cabin during a contaminated air event could not be determined. Therefore, the nature and levels of potentially toxic chemicals entering the cabin during a fume event remain unknown. The number of flights monitored by the Cranfield researchers seems to have been inadequate as fume events are relatively infrequent and do not occur on every flight. The Cranfield study focused on tri-*ortho*-cresyl phosphate (TOCP) and did not specifically measure other isomers of TCP that are more neurotoxic (MOCP and DOCP). The study did not monitor output of contaminants from the auxiliary power unit (APU), which many pilots consider to be a major source of contamination; sampling was undertaken at floor level yet some pollutants exist as a mist dispersed thoughout the cabin; appropriate aviation safety standards do not exist for some of the chemicals detected; and potential synergistic effects of chemical combinations were not considered. The possibility that some individuals may be more susceptible than others to neurotoxic substances (e.g., the very young and old) was not considered. As such, Cranfield's conclusions may be inappropriately reassuring.

5. CONCLUSIONS

The results from this study (and earlier studies), which identify physical and neuropsychological symptoms amongst airline pilots, are disquieting and warrant further investigation, particularly considering that airline pilots are relatively young and usually of sound health because employment selection criteria exclude those with a medical or psychiatric history from entering the profession and working pilots are subjected to regular health checks. Furthermore, pilots are responsible for the safety of hundreds of passengers and any unexplained ill health in a pilot constitutes a flight safety risk.

Given the scientific uncertainty regarding the effects on human health of inhaling pyrolysed engine oil, future research should be commissioned. A large scale epidemiological study is, in our opinion, overdue, but future researchers must pay careful consideration to exposure assessment. Clearly the current study has demonstrated that proxy measures of exposure, such as airframe flown, result in exposure misclassification. Other approaches recommended by the COT involve reliance on self-reported measures of exposure but these are likely to be unreliable due to individual differences in sensory discrimination and the limits of human memory. The COT also suggested researchers try to obtain engineering reports, but these may underestimate the incidence of faulty engine seals as cases may be missed unless ground engineers are notified by cabin crew that fumes have entered the aircraft.

Exposure measures need to be objective and should make use of devices that can reliably detect target contaminants during flight. An alternative approach, which has previously been unavailable to researchers, would be utilization of biomarkers of exposure such as that for TCP exposure developed by researchers from the Universities of Washington and Nebraska. Prospective studies involving them would allow the establishment of dose–response curves and assist in identifying individuals who may be particularly vulnerable to the toxic effects of engine oil fumes.

ACKNOWLEDGMENTS

The authors would like to thank the Independent Pilot Association (IPA) for help with recruitment and the pilots who agreed to take part in this study.

REFERENCES

1. Winder, C. and Balouet, J.C. Aircrew exposure to chemicals in aircraft: symptoms of irritation and toxicity. *J. Occupational Health Safety* (Australia & New Zealand) **17** (2001) 471–483.
2. Winder, C. and Balouet, J.C. The toxicity of commercial jet oils. *Environmental Research* **A 89** (2002) 146–164.
3. Winder, C., Fonteyn, P. and Balouet, J.C. Aerotoxic syndrome: a descriptive epidemiological survey of aircrew exposed to in-cabin airborne contaminants. *J. Occupational Health Safety* (Australia & New Zealand) **18** (2002) 321–338.
4. Winder, C. and Michaelis, S. Aircraft air quality malfunction incidents: causation, regulatory, reporting and rates. Air quality in airplane cabins and similar enclosed spaces. In: *The Handbook of Environmental Chemistry*, vol. 4, part H, pp. 211–228. Springer-Verlag: (2005).
5. Cox, L. and Michaelis, S. A survey of health symptoms in BAe 146 aircrew. *J. Occupational Health Safety* (Australia & New Zealand) **18** (2002) 305–312.
6. Montgomery, M.R., Weir, T., Zieve, F.J. and Anders, M.W. Human intoxication following inhalation exposure to synthetic jet lubricating oil. *Clin. Toxicol.* **11** (1977) 423–426.
7. Coxon, L. Neuropsychological assessment of a group of BAe 146 aircraft crew members exposed to jet engine oil emissions. *J. Occupational Health Safety* (Australia & New Zealand) **18** (2002) 313–319.
8. Somers, M. Assessing over thirty flight crew who have presented as a result of being unwell after exposure to fumes on the BAe 146 jets. In: *Proc. BALPA Air Safety and Cabin Air Quality International Aero Industry Conference*, pp. 131–145. Imperial College, London, 20–21 April 2005.
9. Harper, A. Illness related to cabin air: a survey of symptoms and treatment among commercial pilots and cabin crew. In: *Proc. BALPA Air Safety and Cabin Air Quality International Aero Industry Conference*, pp. 43–51. Imperial College, London, 20–21 April 2005.
10. Heuser, G., Aguilera, O., Heuser, S. and Gordon, R. Clinical evaluation of flight attendants after exposure to fumes in cabin air. In: *Proc. BALPA Air Safety and Cabin Air Quality International Aero Industry Conference*, pp. 107–112. Imperial College, London, 20–21 April 2005.
11. Burdon, J. and Glanville, A.R. Lung injury following hydrocarbon inhalation in BAe 146 aircrew. In: *Proc. BALPA Air Safety and Cabin Air Quality International Aero Industry Conference*, pp. 53–58. Imperial College, London, 20–21 April 2005.
12. Mackenzie Ross, S., Harper, A. and Burdon, J. Ill health following exposure to contaminated aircraft air: psychosomatic disorder or neurological injury? *J. Occupational Health Safety* (Australia & New Zealand) **22** (2006) 521–528.
13. Committee on Toxicity of Chemicals in Food Consumer Products and the Environment. *Statement on the Review of the Cabin Air Environment, Ill Health in Aircraft Crews and the Possible Relationship to Smoke/Fume Events in Aircraft* (2007).
14. Mackenzie Ross, S. Cognitive function following exposure to contaminated air on commercial aircraft: A case series of 27 pilots seen for clinical purposes. *J. Nutritional Environmental Medicine* **17** (2008) 111–126.
15. House of Lords Science and Technology Committee, 1st Report of Session 2007–08. *Air Travel and Health: an Update: Report with Evidence*. London: Stationery Office (2007).
16. House of Lords Science and Technology Committee, 3rd Report of session 2007–08. *Air Travel and Health Update: Government Response.* London: Stationery Office (2008).
17. Liyasova, M., Li, B., Schopfer, L.M., Nachon, F., Masson, P., Furlong, C.E. and Lockridge, O. Exposure to tri-*o*-cresyl phosphate detected in jet airplane passengers. *Toxicol. Appl. Pharmacol.* **256** (2011) 337–347.
18. Schopfer, L.M., Furlong, C.E. and Lockridge, O. Development of diagnostics in the search for an explanation of aerotoxic syndrome. *Anal Biochem.* **404** (2010) 64–74.
19. Marsillach, J., Richter, R.J., Kim, J.H., Stevens, R.C., MacCross, M.J., Tomazela, D., Suzuki, S.M., Schopfer, L.M., Lockridge, O. and Furlong, C.E. Biomarkers of organophosphorus (OP) exposures in humans. *Neurotoxicology* **32** (2011) 656–660.
20. Kidd, J.G. and Langworthy, O.R. Jake paralysis: Paralysis following the ingestion of Jamaica ginger extract adulterated with tri-*ortho*-cresyl phosphate. *Bull. John Hopkins Hospital* **52** (1933) 39–65.
21. Morgan, J.P. The Jamaica ginger paralysis. *J. Am. Medical Association* **248** (1982) 1864–1867.
22. Crump, D., Harrison, P. and Walton, C. *Aircraft Cabin Air Sampling Study; Part 1 and Part 2 of the Final Report*. Institute of Environment and Health, Cranfield University (2011).
23. Van Netten, C. Design of a small personal air monitor and its application in aircraft. *Science Total Environment* **407** (2009) 1206–1210.
24. Snaith, R.P. and Zigmund, A.S. *The Hospital Anxiety and Depression Scale: Manual.* NFER-Nelson (1983).
25. Wechsler D. *Manual for the Wechsler Adult Intelligence Scale-III (UK)*. Psychological Corporation: USA (1998).
26. Coughlan, A.K. and Hollows, S.E. *The Adult Memory and Information Processing Battery: Test Manual*. Leeds: A.K. Coughlan (1985).
27. Spreen, O. and Strauss, E. *A Compendium of Neuropsychological Tests*. Oxford University Press; New York (1991).
28. Trenerry, M.R, Crosson, B., DeBoe, J. and Leber, W.R. *STROOP Neuropsychological Screening Test Manual.* Psychological Assessment Resources. Florida, USA (1988).
29. Borkowski, J.D., Benton, A.L. and Spreen, O. Word fluency and brain damage. *Neuropsychologia* **5** (1967) 135–140.
30. Miller, E.N. *California Computerised Assessment Package Manual (CALCAP)*, 2nd edn. Eric N. Miller and Norland Software (2002).
31. Green, P. *Manual for the Medical Symptom Validity Test.* Edmonton: Green's Publishing (2004).
32. Mackenzie Ross, S.J., Clark, J.S., Harrison,V. and

Abraham, K.M. Cognitive impairment following exposure to organophosphate pesticides: a pilot study. *J. Occupational Health Safety* (Australia & New Zealand) **23** (2007) 133–142.

33. Wycherley, R., Lavender, A., Holttum, S., Crawford, J.R. and Mockler, D. WAIS-III UK: an extension of the UK comparability study. *Br. J. Clin. Psychol.* **44** (2005) 279–288.
34. Schretlen, D.J., Munro, C.A., Anthony, J.C. and Pearlson, G.D. Examining the range of normal intraindividual variability in neuropsychological test performance. *J. Intl Neuropsychol. Soc.* **9** (2003) 864–870.
35. Lezak, M.D, Howieson, D.B. and Loring, D.W. *Neuropsychological Assessment,* 4th edn. Oxford: University Press (2004).
36. Mackenzie Ross, S.J., Brewin, C.R., Curran, H.V., Furlong, C.E., Abraham-Smith, K.M. and Harrison, V. Neuropsychological and Psychiatric functioning in sheep farmers exposed to low levels of organophosphate pesticides. *Neurotoxicol. Teratol.* **32** (2010) 452–459.
37. Hartman, D.E. *Neuropsychological Toxicology: Identification and Assessment of Human Neurotoxic Syndromes*, 2nd edn. New York: Plenum Press (1995).
38. Berent, S. and Alber, J.W. *Neurobehavioural Toxicology: Neurological and Neuropsychological Perspectives. Volume 1: Foundations and Methods.* London and New York: Taylor and Francis (2005).
39. Cook, N. Breathe easy? *RoSPA Occupational Safety Health J.* (2011) 9–13.
40. Learmont, D. *Cranfield University proposes action over toxic cabin air*. http://www.flightglobal.com/news/articles/cranfield-university-proposes-action-over-toxic-cabin-air-364372/